The microgeographies of global finance: High-frequency trading and the construction of information inequality

Matthew Zook
Department of Geography, University of Kentucky, USA

Michael H Grote
Frankfurt School of Finance & Management, Germany

Abstract
Automated high-frequency trading has grown tremendously in the past 20 years and is responsible for about half of all trading activities at stock exchanges worldwide. Geography is central to the rise of high-frequency trading due to a market design of “continuous trading” that allows traders to engage in arbitrage based upon informational advantages built into the socio-technical assemblages that make up current capital markets. Enormous investments have been made in creating transmission technologies and optimizing computer architectures, all in an effort to shave milliseconds of order travel time (or latency) within and between markets. We show that as a result of the built spatial configuration of capital markets, “public” is no longer synonymous with “equal” information. High-frequency trading increases information inequalities between market participants.

Keywords
Geography of finance, high-frequency trading, location of financial trading, information inequality, socializing markets

The geography of capital exchanges
The geography of finance is a diverse field ranging from the identification of global economic centers to analyzing how financial practices differ between regions to investigating how investment contributes to knowledge exchange (Clark, 2002; Sassen, 1991). Common to all this work is a recognition that despite its mobility and fungibility, capital is embedded in society and possesses distinct spatialities that govern its operations and shapes the procedures and institutions of the financial industry. In line with this previous work, this...
paper focuses on a particular practice with finance – high-frequency trading (HFT) in capital markets – and asks a series of questions, namely: How do the current designs of global exchanges leverage differences in spatial location? How do traders incorporate this in their trading strategies? What are the benefits and risks associated with these strategies and how are they distributed?

In answering these questions, we provide a detailed study of socio-technical assemblages that make up HFT and show that capital markets are fundamentally shaped by geographical concerns in the form of reducing latency between market events. Moreover, HFT transmits a number of potentially onerous effects to the larger society such as major, albeit short-lived, downturns in capital markets and many deviations from fair prices in stock trading. Importantly, these increased risks in the economy are not restricted to capital market participants but extend to society at large. Additionally, we show that the very core of HFT is the creation of information asymmetries. This runs counter to the general societal expectation that everyone should have equal access to market information, and important information should therefore be public. HFT separates “equal” from “public” information by getting and using information in tiny moments – measured in micro- and nanoseconds – before everybody else. This has consequences for capital markets that rely on the aggregation of information and preferences from all participants to form the most informative prices. When capital markets fail to do so, prices will deviate from their fair level with significant implications. For example, in the medium term, capital markets might be compromised in their function when potential participants lose trust in a fair pricing process. This is why regulators developed sophisticated rules to block insider trading; e.g., people with superior private information are banned from trading or using this information. In a related manner, HFT aims to privatize public information for just long enough to profit and in so doing public and equal are no longer the same. Thus, we argue that HFT is an equity issue and society – understood as all non-high-frequency traders whether they participate in capital markets or not – should be protected from unnecessary risk and unequal access to information.

As long as current market designs are based on non-socialized understandings of markets – abstracted arenas for buy and sell orders – and geography – as merely a physical space to traverse – the information inequalities designed to operate within capital markets and the risks that are transmitted to all regardless of participation in capital market trading will not be acknowledged. Instead, we call for re-evaluation of the assumption and rules by which exchanges operate that currently distribute of information, benefits and risks on behalf of certain traders rather than market participants and society more generally.

The spaces of market information

Much of the geographical research on capital markets has focused on explaining the locations of global financial centers such as London or New York. In addition to longstanding arguments related to agglomeration economies, researchers highlight the ways that the location of traders provides advantages; particularly as exchanges transitioned from physical trading floors to electronic trading (Zaloom, 2006). Examples include Lo and Grote’s (2003) work on the benefits of proximity to other traders and changing views of the strengths and weaknesses of companies. In a similar vein, Hau (2001) finds proximity to the headquarters of firms makes trades more profitable than others located elsewhere and Clark and Monk (2013) highlight how firms’ decisions to produce and consume information influence their success.

Thus, economic geographers characterize capital markets as geographically embedded institutions (Muellerleile, 2013) due in large part to information and knowledge which has its own distinct spatialities (e.g., Wójcik, 2007, 2009). Information about world events;
deep knowledge about the prospects of firms; information about commodity prices and the actions of other participants: each data point contributes to buy, hold and sell decisions. The market value of privileged access to information has long been recognized and accounts for prohibitions against front running.

Minimizing informational differences, however, has historically been difficult given the materiality of exchanges. Physical presence on a trading floor provides better information on market reactions relative to those outside and physical attributes such as brightly colored trading jackets or being able to “run faster across the trading floor or shout louder than their counterparts” (Gomber et al., 2011: 10) provide direct advantage in trading. Ironically, the transition from trading pits to electronic exchanges which extended the geographical range of traders’ access markets has been accompanied by a modulating significance, but not irrelevance, of spatial proximity. While traders were free to participate from all places there has been an intensity in financial agglomerations, particularly in the case of London as traders seek an informational edge (see Engelen and Grote, 2009; Grote, 2007). Not only is co-presence important for human traders but distance between the technological artifacts of matching engines and computer circuits have proven important actants (Latour, 2005) in the construction of competitive advantage in today’s markets.

The creation of markets and the role of technologies

The centrality of information means that technologies which compress time-space (Harvey, 1989) have played important roles throughout the history of organized stock markets. Nathan Meyer Rothschild is said to have used carrier pigeons to trade in London on the information of Napoleon’s historic defeat in Waterloo in 1815, and traders in stock exchanges are regularly among the earliest users of modern technologies – be it the telegraph in the 1850s, the telephone in the 1880s, or the screen traders in the 1980s (see Leinweber, 2009). High-frequency traders’ investments in new fiber optics routes and computers represent the latest iteration in this trend as traders seek to maximize competitive advantage in capital markets. The introduction of new technologies, however, also raises questions about how capital markets are recreated by actors embedded in the larger society. While neoclassical economics conceptualize markets as pure rational competitive environments, a wide range heterodox economic approaches – emerging from economic sociology and geography – emphasize the social dimension of markets and economic functions. Building upon the earlier work of Polanyi (1957) and Granovetter (1985), scholars such as Storper and Harrison (1991) and Saxenian (1994) applied institutionalist approaches to analyze the competitive landscape of firms in particular regions. This emphasis on more than market factors is central to the cultural turn within economic geography (Thriff and Olds, 1996) and its focus on the importance of social and cultural factors within economic processes. This emphasis on culture in the economic continued (Barnes, 2003; Thriff, 2000) with Castree (2004: 206) arguing that “there are no such things as economy and culture ontologically speaking”. The result was an economic geography that seeks accounts “that are open-ended, reflexive, and self-consciously perspectival and partial” (Barnes, 2001: 561).

Parallel to this shift within economic geography, actor network theory (ANT) emerged from social and technology studies. Scholars within the ANT tradition highlight all factors within a social process – both human action as well as those previously disregarded. Within ANT, cultural factors, objects, regulations and technologies are all analyzed for their contribution to the performance of a particular social process via webs or actor-networks. This inclusion of technology is particularly useful for highly technologically mediated activities such as HFT. As Latour (2005: 65) argues, the social
...is an association between entities which are in no way recognizable as being social in the ordinary manner, except during the brief moment when they are reshuffled together [...] Thus, social, for ANT, is the name of a type of momentary association which is characterized by the way it gathers together into new shapes.

In practice, ANT integrates both the human and non-human actants (a preferred term since actor is so closely associated with humans) into a common analytical framework that allows all actants to exercise agency. ANT is particularly well suited to analyze the practices of HFT given the complex interplay between human strategies and technological infrastructure in this field.

We follow Çalışkan and Callon (2010) who apply the ideas of ANT to study what they refer to as marketization or the process through which the behaviors and institutions associated with markets are established. Using the methodological construct of socio-technical assemblages they highlight the performative role of knowledge – both economic theories but also tactical field-based knowledge – as it is deployed by actants in market creation. Çalışkan and Callon (2010: 16, 23) also highlight the role of what they call “material technologies” in structuring markets and facilitating experimentation thus emphasizing the techno-social aspects of marketization. This approach, with its concern for the dynamic range of actants and assemblages, risks limiting analysis to a descriptive narrative that does not connect these changes to larger political economy and institutional imperatives. In short, ANT can produce deep understandings of the shape of evolving processes but remains largely silent on issues of how these changes affect the societal space outside the networks.

But it need not be so. MacKenzie’s (2006) study of the linkage between economic models and the design of financial markets – derivatives, futures and options – offers a good example of an approach that links ANT approaches to larger concerns. MacKenzie traces the history of the Black-Scholes-Merton algorithm and argues that it was adopted not for its accuracy but for its ability to solve a variety of political and operational problems faced by the finance industry. In a later work, Millo and MacKenzie (2008: 31) emphasize the model’s key role in pricing – “a stream of methodologically valid information (although not always realistically valid)” – which helped legitimize these markets, a key goal of those seeking to profit. Zeroing in to the specifics of the information technology actants they argue that “inhuman speed and efficiency were the factors that kept the “facts machine” of financial risk management running smoothly” making the connections of this market a “techno-social network rather then [sic] only a social one.” In a similar manner, our approach uses ANT to explicate the techno-social assemblages fundamental to the operation of HFT while simultaneously highlighting the ways these assemblages are embedded in capital exchanges grounded in societies and regulations which shape they ways wealth and risks are distributed.

The social basis of HFT

Our first goal in this paper is to outline the socio-technical assemblages that make up current capital markets and explicate how these extend across space. Beginning in the 1980s, trading shifted from physical territories, as “algorithmic trading” emerged (see Leinweber, 2009), heralding the declining relevance of the trading pit. With algorithmic trading assemblages, stock brokers place orders via computer terminals connected to the central computer of an exchange (or matching engine). This actant matches buy and sell orders in milliseconds
without direct human agency, although actual trading strategies remain the purview of human traders acting in the interest of their firm or client within a particular regulatory environment. Algorithmic trading helps reduce transactions costs and helps minimize impacts on market prices, a particularly important feature for large investors as well as exchanges seeking to minimize transaction costs.

This integration between human actants who were devising trading strategies and computer actants conducting the actual trading deepened as algorithmic trading evolved into complex and reactive software systems. This ultimately led to HFT systems in which computer platforms and proprietary software actants automatically pursue a range of trading strategies devised and coded by human traders. For example, exploiting price differences between exchanges by exchanging information across hundreds of kilometers a few milliseconds faster than competitors. These HFT strategies marked an important shift in market design and activity making geography a fundamental variable in global capital markets even as the same technology enabled distantiated interactivity (Cetina and Bruegger, 2002). This is a point emphasized by Wójcik (2011: 147) who notes that capital markets have not “dissolve[d] in virtual space”, highlighting the continuing importance of physical proximity, albeit within server farm rather than trading pits.

Given this shift in markets design, high-frequency assemblages focus on a different set of issues than other market participants. Easley et al. (2012) highlight the dissimilarities thusly:

...low frequency traders (LFT) converse on subjects as broad and complex as monetary policy, asset allocation, stock valuations, financial statement analysis, and the like. HFT Conferences are reunions where computer scientists meet to discuss TCP/IP connections, machine learning, numeric algorithms to determine the position of an order in a queue, the newest low-latency co-location architecture, game theory, and most important of all, the latest variations to exchanges’ matching engines. One would conclude, correctly, that the LFTs and the HFTs are seemingly worlds apart. (Easley et al., 2012: 21)

MacKenzie et al. (2012) list a number of HFT trading strategies including market-making, statistical arbitrage, arbitrage between geographically separate markets, order-anticipation, and the illegal but not unknown practice of momentum ignition. The first – market making or the simultaneous entry of bid and ask orders with a profitable bid-ask spread – is also the one most discussed by proponents of HFT as it generates market liquidity. Liquidity, or the ability for traders to sell shares when they wish, is key for smoothly operating markets. Exchange operators often provide incentives (or rebates) to market makers which can be a considerable source of income for some HFTs (see MacKenzie et al., 2012).

The remaining strategies of HFTs are much less positive and are based on informational inequalities enabled by multiple trading venues, many of relatively recent origin. These alternative trading venues have attracted considerable liquidity from the incumbent stock exchanges: In 2005, about 80 percent of the trading in stocks listed on the New York Stock Exchange (NYSE) took place there while today it is closer to 25 percent (Menkveld, 2013) with Europe developing in a similar pattern. These market venues, also referred to as Multilateral Trading Facilities (MTFs), include various design tweaks – most famously anonymous trading – which led to the label “dark pools” of money. MTFs are useful to HFT because several markets trading the same or very similar securities means that slight differences in price between stocks or market provides opportunity for statistical and geographic arbitrage. Traders seeking to profit from these opportunities exploit and create information inequality in the assemblages acting at the level of electronic circuits and milliseconds, far removed from the metrics associated with traditional investors.
The design of space-based strategies of HFT

To understand the complex interplay between market providers, technology and different sorts of market actants, we use an in-depth study of Deutsche Börse, the leading German stock exchange. The experience of the Deutsche Börse is similar to other large stock exchanges in terms of its sophisticated technology – with some noted exceptions – and market position; a point stressed by our interview subjects. In addition to five formal interviews with human traders and Deutsche Börse’s employees in 2013 and 2014, we also drew insight from a much larger number of informal talks with bankers, traders, IT-providers, finance professors, and regulators during the same time period. This is made possible, inter alia, because one of the authors is affiliated with an educational institution closely affiliated with bankers and traders. The gathered information has been cross-checked and verified by academic research publications within finance, industry reports, technical specifications and evaluations, advertising and public relations materials, the latter primarily from Deutsche Börse.

HFT is remaking markets through continual refinement of algorithmic actants designed to maximize speed (or lower latency) between any two market actions (see Budish et al., 2013). Most matching engines at capital exchanges facilitate continuous trading by fulfilling orders first by price, highest buy orders or lowest sell orders are executed first, and second by time, orders with the same limit price are executed in the order they arrive.

Although the trading process appears “continuous” for human actants, it is in fact not the case; as there are still discrete points in time when trades are settled. Budish et al. (2013) show that these – extremely small – differences in execution time create “purely technical arbitrage opportunities” (2) that can be exploited by HFT. For example, an increase in the value of one of two highly correlated financial instruments traded on an exchange means that it is highly probable that the price of the second security will rise as well. Traders using this strategy rely upon coded actants with privileged access to information that allows for the microsecond reactions necessary to benefit from the momentarily lower price. Leveraging information about price differences between locations – referred to as geographic arbitrage – represents a similar strategy of constructing informational advantage albeit via long-haul cable routings.

To illustrate some of the effects of HFT assemblages, we “follow” a large algorithmic (albeit not high frequency) buy order as it hits different and physically separated markets. This example comes from a report by LiquidMetrix (2013), a financial market data research firm, and is typical example of this process (Toulson, 2013). A buy order for a Swedish stock is issued by a human trader in London in concert with “smart order routing” (SOR) software which divides large orders between different stock exchanges to get the best price. The order was split into four parts, and sent simultaneously to three London-based trading venues and the Stockholm exchange and executed in 58 milliseconds via 70 smaller trades or fills. The buy order first hits the London markets and is filled based on existing sell orders but the microscopic difference in time it takes for the order to reach the Swedish exchange results in a different outcome.

During the transit to the Stockholm exchange, HFT algorithms acting on behalf of unknown human traders learned of the buys in the London markets and cancelled 7 out of 11 of their sell-offers in the order book in Stockholm. Some of the other sell-offers were hit (“stolen”) by even faster HFT programs and executed in front of the original large order, presumably to resell to the original buy order at a higher price. LiquidMetrix concludes:

This implies that the HFTs were able to react to the executions on London and send messages to cancel orders in XSTO [the Stockholm stock exchange] faster than time taken for the SOR
orders to reach Stockholm. Other HFT market participants (not necessarily the same firms) aggressively traded ‘in front’ of the SOR slice. (Toulson, 2013: 39)

After being “surprised” in London by the large order, HFT assemblages leveraged their advantage in compressing time-space to adjust their bids in other locations, ultimately resulting in a higher price for the original buyer. While the process took place in the micro techno-social assemblages of the HFT, the results (i.e., changes in prices) were experienced also by human traders outside these networks.

Other strategies used in HFT include “quote stuffing” or issuing a large amount of orders that are immediately canceled (see Tse et al., 2012a). The order and cancellation increase traffic at matching engines, slowing competitors reactions as well as changing the exchange’s price allowing the perpetrator to profit from arbitrage at other markets, particularly dark pools (Tse et al., 2012b). According to an interview at Deutsche Börse, orders can be mere microseconds apart, thereby allowing for more than 100,000 orders per second from a single market participant. European stock exchanges have reacted to “quote stuffing” by charging penalty fees when the ratio of orders to trades becomes too lopsided, defined in the case of the Deutsche Börse as more than 2500 orders to one trade (see Tse et al., 2012a).

All the techniques discussed so far are legal but the human traders with HFTs assemblages have deployed a number of greyzone and illegal strategies as well (see Easley et al., 2012; Mackenzie et al., 2012) including “momentum ignition”, which tries to trigger other market participants to trade a stock in a certain direction and then profit from later price reversals (Tse et al., 2012b). Another technique is “layering”, or placing a number of orders to induce the impression of buy or sell pressure in the market and move prices accordingly (see Kirilenko and Lo, 2013). While these types of practices are not unknown in slower trading environments, the speed at which markets operates – exchanges timestamp transactions in terms one billionth of a second (see Eholzer, 2013) – means illegal manipulation literally occurs in the blink of an eye complicating detection of this illegal profit-taking and punishment of the human traders instigating it. An investigation by the US Securities and Exchange Commission notes:

That day, at 11:08:55.152 a.m., the trader placed an order to sell 1,000 GWW shares at $101.34 per share. Prior to the trader placing the order, the inside bid was $101.27 and the inside ask was $101.37. The trader’s sell order moved the inside ask to $101.34. From 11:08:55.164 a.m. to 11:08:55.323 a.m., the trader placed eleven orders offering to buy a total of 2,600 GWW shares at successively increasing prices from $101.29 to $101.33. During this time, the inside bid rose from $101.27 to $101.33, and the trader sold all 1,000 shares she offered to sell for $101.34 per share, completing the execution at 11:08:55.333. At 11:08:55.932, less than a second after the trader placed the initial buy order, the trader cancelled all open buy orders. At 11:08:55.991, once the trader had cancelled all of her open buy orders, the inside bid reverted to $101.27 and the inside ask reverted to $101.37. (Securities and Exchange Commission, 2012)

A few seconds later – from 11:09:00.881 to 11:09:01.792 – a similar series of orders were placed, albeit from opposite direction as the human trader of the HFT assemblage directed algorithmic actants to shift price yet again by providing an impression of market demand (Securities and Exchange Commission, 2012). It is also important to note that the time stamp in this example is demarcated in microseconds, e.g., it took 160 milliseconds to place 11 orders and the overall attempt to layer the market took less than one second. While clearly the design of a human trader to manipulate prices, this strategy is only successful in concert with algorithmic actants which also makes geography a key variable in the design of capital markets.
**Geography of HFT**

The spatiality of capital exchanges is based upon the geographies of two distinct types of information: First is data from the world outside the market, either macroeconomic (e.g., consumer confidence surveys), or firm-specific (e.g., quarterly reports) used by analysts in firm and industry valuation. The second source of information derives from the market itself including pricing or large orders (Toulson, 2013). High-frequency traders primarily – though not exclusively – rely upon the latter because reacting to market information requires little direct human agency. Thus, lowering latency in information exchange becomes the paramount concern.

**Long haul routes and constructing market advantage**

Geographical arbitrage – profiting on price differentials between exchanges for the same or similar securities – is an important ingredient for many HFT strategies. Two key determinants of latency\(^2\) are the physical distance between markets (including any non-direct routing) and the propagation delay of the medium used; fiber optic cables transmit information at approximately 70 percent the speed of light while satellite or microwaves approach the speed of light. This means that information sent by fiber optic cable between New York and London takes approximately 8 microseconds longer than if sent by microwave (o3bnetworks, n.d). Historically such small differences in latency have only concerned computer network engineers but HFT has made it central a central element to the design of capital exchanges.

Shortening fiber optic cable routes has been a key part of these designs: A new cable connection between the UK and the US (costing $300 million) reduced latency by 6 milliseconds while an unknown amount has been spent on connections between the UK and Japan to reduce one-way latency by 60 milliseconds (Anthony, 2012). Investment in low latency routes has been strong in the US, particularly between its two main stock exchanges, Chicago and New York accomplished in part by drilling small tunnels through mountains (MacKenzie et al., 2012). The project has been estimated to cost $300 million and has reduced the round-trip latency between New York and Chicago to 13.33 milliseconds (Spread Networks, 2010). This speed improvement of 3 milliseconds over existing networks – which followed more circuitous paths – allowed its owners to charge HFTs up to 10 times more than incumbent slower infrastructure providers (Laughlin et al., 2013).

This faster fiber connection was followed by other efforts, and Laughlin et al. (2013) estimate that 15 to 20 private microwave networks between New York and Chicago have been installed, each costing approximately $8 million and reducing latency by 5 milliseconds given microwave’s lower propagation delay relative to fiber. The shortest connection between the two trading centers is already crowded by microwave stations (MacKenzie, 2014). Indeed, efforts to compress time-space are approaching physical limits as current latencies approach the actual speed of light and even signals “beamed directly through the Earth [...i.e., cutting short the curvature of the earth...] would cut only \(~1\) km or \(~3\)\(\mu\)s [milliseconds] from this latency” (Laughlin et al., 2013: 8, footnote 7). Europe has seen similar developments with a microwave connection between London and Frankfurt installed in 2012 reducing round-trip latency from about 8.35 milliseconds to about 4.6 milliseconds in part by avoiding public and potentially congested routers (Anthony, 2012). The competitive advantage of these low latency assemblages that accrue to market participants is aptly demonstrated by its operator’s advertising that “Microwave technologies are the ideal scenario for trading ahead of the market – not for the entire market, but for a select group of assets with a clear strategy for high precision and high speed” (Watkins, 2013).
The “selectiveness” of the constructed information inequality is further demonstrated by the fact that its existence was initially not made public (Onstad, 2013), presumably providing its users with a valuable advantage.

**Measuring meters at co-location facilities**

While the regions connected by long-haul fiber and microwave routes have remained reasonably consistent over the past decades (Grote et al., 2002) the relevant places within each region have changed. In contrast to historical practice when stock exchanges exerted a very localizing force – traders needed to be on the floor (see Grote, 2007) – HFT has shifted the relevant locality at which efforts to maximize proximity are focused. Rather than the trading pit – largely symbolic relics of exchanges today – the center of markets today are “co-location facilities” or server farms where the computer actants of HFT assemblages share co-presence with the “matching engine” of the exchange which conducts the actual trades. In HFT face to face proximity is no longer key, and human actants operate out of disparately located offices and rely on strategies of virtual sociability when needed (Cetina and Bruegger, 2002) – but computer colocation is absolutely essential.

This has shifted the locus of actual trading from city centers to the peripheries of financial regions with the matching engines of the NYSE and NASDAQ located in Mahwah and Carteret, New Jersey respectively 55 and 35 kilometers from Wall Street while the co-location facility for the Chicago Mercantile Exchange is 66 kilometers from city center in in Aurora, Illinois. European exchanges exhibit a similar pattern with the servers of the NYSE Euronext group located in Basildon, 55 kilometers away from London and the co-location facilities for the Deutsche Börse based in the outskirts in Frankfurt rather than the financial district.

These co-location facilities are the new trading centers, where orders from both high- and low-speed traders are matched and executed, and location within the facility is key concern. HFT wants to minimize distance to both the exchange’s matching engine and the server providing market information on trades and when microseconds (a millionth of a second) can spell the difference between a favorable trade or not, meters matter. According to an employee of Deutsche Börse, as recently as 2010, HFTs would pay extra for the location closest to the stock exchange computer within the same room. Due to complaints from other HFTs, latency differences – on the order of ten microseconds or one ten thousandth of a second – stock exchanges including the Deutsche Börse now offer standardized cable lengths within the co-location centers (MacKenzie et al., 2012). The importance of these internal geographies of co-location facilities is reflected in marketing materials: the Deutsche Börse (2011) notes that their “network also offers a standardised fibre length” providing “all market participants the same low latency” and Eurex (2013) assures its customers that “The physical location of the high-frequency Eurex Enhanced Transaction Interface gateways in the Equinix data centre relative to the room where a matching engine resides has no impact to the order latency”.

Even with standardized cable lengths, the human practitioners of HFT continue to manipulate co-location geographies to their advantage. According to one interview in Deutsche Börse, placing a HFT server directly below the exit nozzle of the air conditioning system provides a somewhat cooler operating temperature, therefore allowing it to run faster and thus react to market information at a higher speed than its not-so-cool peers. Again in reaction, the exchange operators of co-location facilities now provide equal access to the air conditioning system for all servers in the room although the same interview report that
HFTs have experimented with the number and spacing of server within racks to optimize temperature and computing performance.

Maximizing speed within the computer

This focus of gaining every speed advantage does not end at the colocation facility but extends to the level of electronic circuit boards as HFTs seek to reduce latency within actants at even the most micro scales. As noted in an IT industry trade group report:

In the online and Co-Lo [i.e., co-location] based financial trading markets, performance, both in terms of latency and throughput is paramount. It is the difference between a firm being „in the market“ or not. Complete trading systems are built from many complex elements, including market data capture, trading algorithms, trade execution, and inflow risk analysis. These elements run on critical infrastructure components, hardware, software, network and connectivity all of which must interoperate with each other. (ONX Consortium, 2012: 2)

Most HFT servers and operating systems are custom designed to maximize the speed of processing information and differ in important ways from general purpose computers. A key part of the internal geography of non-HFT computers is to ensure that different applications don’t conflict and a normal operating system divides processing power according to standard protocols to allow for multitasking. There are any number of key interface points – the operating system to the network card, the network card to the local area network (LAN) – at which data is passed and processed in an orderly and robust fashion. This standardization, however, comes at the cost of minuscule delays that loom large in HFT. For example, advertising from developers of high-performance computing, reminds clients that they are operating in “a world where every nanosecond counts” and boast that they offer “a new class of application accelerator […] which moves application processing into the network adapter for remarkably fast, on-the-fly processing of network data […] for ultra-low latency, high message rate performance, and seamless application compatibility” (Solarflare, n.d).

In practice, low-latency computing is maximized in a number of ways such as the use of Field Programmable Gate Arrays (FPGA), computer circuits that are directly configurable after installation and allow traders to bypass an array of “slower” software protocols (Leber et al., 2011; Lockwood et al., 2012). This represents a microscopic spatial strategy on a server’s network interface card bringing a trader’s computer “closest to the wire” of an exchange’s matching engine (HFT Review, 2012). Another technique for low latency is designing trading software with a “kernel bypass” so that the coded HFT actants avoid the server’s operating system – busily integrating hardware and software of the computer – to achieve the most time-space compression possible (see Eholzer, 2013). Given the extremes to which HFT strategies go, precise time metrics become very important, e.g. to evaluate one’s trading strategies. Exchanges such as the Deutsche Börse offer super-exact timing with “accuracy […] in the sub-microsecond range” (Eholzer, 2013) and come at a considerable cost for market participants (Eurex, 2012).

Whether it be laying new fiber optic cable or optimizing network card performance, the human traders directing HFT assemblages are profoundly preoccupied by geographic concerns of location, travel time and proximity. While this has enabled some to make vast fortunes, the design of capital exchanges, particularly increasing information inequality among market participants, has been a key means through which HFT traders have extracted wealth.
The risks of HFT

In a report commissioned by Deutsche Börse, Gomber et al. (2011) conclude that HFT strategies provide a number of benefits including higher liquidity making it easier to exchange securities without affecting price, better price discovery leading to the “right” price and higher market efficiency with similar securities trading for similar prices. In short, the report argues that HFT is a net benefit for exchanges and further asserts that “[a]cademic literature mostly shows positive effects of HFT based strategies on market quality” (Gomber et al., 2011: 2). While not disputing some of the individual effects of HFT (on liquidity for example), the overall conclusion of higher quality markets only takes into account an overly limited definition of market quality or the social dimensions of market creation. At this point, our analysis also shifts from a focus on the socio-technical assemblages of HFT to the ways in which these assemblages allow human traders embedded in markets to capture profit and distribute risk. The strongest connection between the larger societal space outside the HFT-assemblage is via those risks. Many market participants, but – more importantly – people that do not take part in capital market trading are exposed to risks stemming from HFT. Large interruptions on capital markets brought about by HFT might lead directly to the destruction of wealth and indirectly to a chain reaction resulting in an economic downturn. On a smaller scale, firms that experience increased volatility in stock prices might face a higher perception of their risks and thus higher financing costs.

The prevalence of HFT in markets has raised concerns of the public (see Lewis, 2014) and regulators worldwide. For example, Germany passed a “High-Frequency Trading Act” in May 2013, ahead of an expected directive from the EU, to prevent “extreme, irrational price fluctuations that are completely disconnected from developments in the real economy” (Bundesfinanzministerium, 2012). The worries are not unwarranted; Johnson et al. (2013) have documented 18,520 “ultrafast extreme events” in stock market pricing within the US between January 2006 and February 2011 in which assemblages engaged in behavior not anticipated by those who run the markets. The most well-known example is the infamous “Flash Crash” on 6 May 2010. In the early afternoon, the Dow Jones Industrial Average index declined about 7 percent in a few minutes, only to bounce back in less than half an hour, to approximately the level before the drop (see Figure 1). During this time, some shares dropped to prices of 1 cent while Apple stock briefly traded at $100,000 per share.

![The “Flash Crash,” May 6, 2010](image)

**Figure 1.** Charting the flash crash, 6 May 2010.
Kirilenko and Lo, 2013). While there is no consensus about the exact causes and structure of the crash, its HFT is generally deemed a key player in amplifying the shocks in both directions (Kirilenko et al., 2011).

Facebook’s initial public offering via NASDAQ on 18 May 2012 offers another example of HFT impacting the function of markets. Due to very heavy demand from HFT, NASDAQ’s computers took about 5 milliseconds to compute the opening price – instead of the usual 0.04 milliseconds. Since NASDAQ allows traders to change orders before the final price is made public, the algorithm actants of many HFT assemblages ordered and cancelled orders during the “delay”. This caused the software to recalculate the opening price, increasing the delay even further, resulting in an extremely rocky start. It took considerable time to sort things out and “many customer orders […] went unfilled for hours or were never filled at all, while other customers ended up buying more shares than they had intended” (Kirilenko and Lo, 2013). This represents a moment in which the actions of HFT assemblages are transmitted to the society at large. Still another example is a programming error by a human technician that resulted in Knight Capital’s trading algorithm pushing high price movement in 150 stocks on 2 August 2012 resulting in a loss of more than $450 million for the company (Kirilenko and Lo, 2013).

More alarming is the vulnerability of capital markets to direct manipulation by actants who are generally not considered to be part of the market but able to insert themselves into the larger trading assemblage. On 23 April 2013, the twitter account of the Associated Press (AP) news agency was hacked and sent out a – completely false – message reporting “Two explosions in the White House and Barack Obama is injured”. The AP twitter account had about two million followers, including HFT algorithms using feeds to scan for news with potential price impacts (Gomber et al., 2011: 24) and upon seeing these tweets, these actants reacted quickly. The main stock market indices plunged immediately about one percent – equaling a loss in value of at least $136 billion – before recovering a few minutes later (Domm, 2013). While a hacker group associated with the Syrian government claimed responsibility for the message, our interviews (as well as a review of HFT internet forums) reveal that many human traders are distrustful of this explanation. After all, this situation was a perfect opportunity for informed traders to profit which challenges the legitimacy of techno-social process of pricing in markets (Millo and MacKenzie, 2009). While there is no evidence that this was the case, it highlights how HFT can undermine confidence in the market.

Indeed, we argue that the risks associated with HFT are quite large: Whole markets might drop and malfunction due to computational interpretation, lack of capacity, programming mistakes or targeted manipulation. At the level at which coded actants conduct actual trades, there is little time or space for a human sanity check. So far, markets with HFT-related crashes have quickly returned to previous levels but a report on the Flash Crash for the UK Government outlines how a massive swing in one market might propagate throughout the global economy:

Specifically, we think that the true nightmare scenario would have been if the crash’s 600 point down-spike, the trillion-dollar write-off, had occurred immediately before market close: that is, if the markets had closed just after the steep drop, before the equally fast recovery had a chance to start. Faced with New York showing its biggest ever one-day drop in the final 15 minutes before close of business on May 6th, and in the absence of any plausible public-domain reason for that happening, combined with the growing nervousness that the Greek government would default on its sovereign debt and throw the entire Eurozone economic union into chaos, traders in Tokyo would have had only one rational reaction: sell. The likelihood is that Tokyo would
have seen one of its biggest ever one-day losses. Following this, as the mainland European bourses and the London markets opened on the morning of May 7th, seeing the unprecedented sell-offs that had afflicted first New York and then Tokyo, European markets would have followed into precipitous freefall. None of this would have been particularly useful in strengthening confidence in the Greek debt crisis or the future of the Euro, either. And, as far as we can tell, the only reason that this sequence of events was not triggered was down to mere lucky timing. Put simply, on the afternoon of May 6th 2010, the world’s financial system dodged a bullet. (Cliff and Northrop, 2010: 7)

The effects of future socio-technical assemblages of ANT could be more severe and long-lasting, especially if trust in the fairness of markets or macro-economy in shaken; or when a multitude of secret HFT strategies reinforce one another to perpetuate declines that do not halt after a couple of minutes or hours. In short, current market designs are flawed in that much of the risk profile associated with HFT is still unknown:

It’s natural to have the model retrain itself each night based on each day’s data. Humans might not know what incremental modifications today’s data introduced in this adaptive process. If everybody’s doing this, you can easily imagine various effects. That could be a bad thing from a stability standpoint in the markets. […] The quant meltdown [another incident involving quantitative trading strategies] demonstrated the unexpectedly high correlation between quant strategies who believed they were doing different stuff from each other. Under ordinary circumstances these correlations weren’t so apparent. When we decide to exit our positions at the same time, we’re going to painfully realize our correlation. (Michael Kearns cited by Salmon, 2011)

In short, given the evidence of known and unknown risk associated with HFT we argue that current market designs are in need of change, particularly, since current structures go hand in hand with increased information inequality among market participants.

**HFT and information inequality**

HFT fundamentally depends upon constructed information inequity in order to trade in ways that bear some similarity to the illegal practice of front running. In pre-computerized trading, front-running entailed receiving notice of a large order before it hit the stock market and then trading the target stock before the larger order was executed. This would move the share price (to the disadvantage of the first buyer) and allow the front runner to profit. As outlined in the example of a buy order for a Swedish stock highlighted previously, a similar tactic is used by HFTs to profit on the basis of unequal access to information constructed via their space based strategies.

In addition to market information, high-frequency traders also engage in front running-like activity by buying special relationships with sources of macro-economic information. For example, the news agency Thomson Reuters distributes a “consumer confidence” indicator compiled by the University of Michigan which is known to drive markets. The indicator becomes public information when it is distributed freely at 10 am but paying clients of Thomson Reuters receive access via a conference call five minutes earlier and high-frequency traders have negotiated even more privileged access.

... the contract carves out an even more elite group of clients, who subscribe to the “ultra-low latency distribution platform,” or high-speed data feed, offered by Thomson Reuters. Those most elite clients receive the information in a specialized format tailor-made for computer-driven algorithmic trading at 9:54:58.000, according to the terms of the contract. (Javers, 2013)
Any investor relying solely on the public release of the consumer confidence data would be behind those with access to the conference call by five minutes (not including the time to actually analyze the meaning of the new figures). Even investors on the conference call are already late by two seconds, an enormity of time for HFT algorithms (NANEX, 2012). The consumer confidence indicator is but one of many similarly constructed opportunities for HFT to leverage information inequality (Durden, 2013; Mullins and Barrett, 2013).

These kinds of tactics are not limited to relations with news agencies but are integral to the way capital markets cater to high-frequency traders. To pick but one example, Alpha Flash, a news feed product of Deutsche Börse is marketed in the following way:

AlphaFlash®, Deutsche Börse’s low latency algo news feed, delivers global market moving macroeconomic indicators and corporate news. It was developed for the most demanding algo traders at hedge funds, trading firms and financial institutions using resources from our fully accredited news agency MNI, whose journalists have direct access to government lock-up rooms and embargoed releases. [...] The Chicago Report disseminates the ‘Chicago PMI’ [purchasing manager index] three minutes before public release. The Fitch Ratings Feed delivers low latency Fitch rating announcements. (Deutsche Börse, 2013)

The advertising copy for Alpha Flash highlights its access to strategic information points, “government lock-up rooms”, as well as its ability to get key indicator data prior to public release. Clearly, the Deutsche Börse is marketing this service as a way for HFT to gain an informational advantage over other investors. To place this in context, German stock market regulations (and most others) explicitly require firms traded on exchange to distribute information to market participants equally to ensure that investors will not be at a disadvantage vis-à-vis insiders (see Hasbrouck and Saar, 2013). The logic behind this rule is outlined in the introduction to the German Securities Trading Act by the German capital market regulator BaFin:

Only if listed companies notify all market participants rapidly and comprehensively of any insider information can investors make well-founded decisions and not be put at a disadvantage to insiders. For that reason issuers have a statutory obligation to disclose immediately [...] facts about their company that are not public knowledge if such information has the potential to influence the price of the financial instrument [...]. For the purposes of such disclosure, under the terms of the Act an issuer in Germany must use a widely disseminated electronic information dissemination system and a combination of media to disseminate the information Europe-wide. (BaFin, n.d)

The intent of this law is to eliminate insider trading and ensure information equality among traders. Thus, it is ironic that once the information is released (or made public), the entirety of the business model for low latency news services is to construct information inequality for a special class of investors. As a result, “public” is no longer synonymous with “equal” precisely because the actants within markets and the speed at which they operate has been so radically transformed. The seconds or minutes it takes a human trader to absorb and evaluate new information is an eternity to the HFT algorithms.

Ensuring equal access to information has long been a key preoccupation and task for regulators (see, e.g., Securities and Exchange Commission, n.d). It remains unclear whether this behavior of preferential treatment for one group of traders over another is acceptable by the authorities. For example, the SEC in the US is currently investigating Thomson Reuters due to hefty trading activity occurring 15 milliseconds before data on manufacturing in the US had been officially released. The worse than anticipated data coincided with 30,000 sell orders of futures contracts in more than 300 different shares – in the space 15 milliseconds
(Van Steenwyk, 2013), presumably represent an effort of some traders to profit from their information advantage constructed via the peculiar spatial structure of capital exchanges.

Conclusion
As outlined above, the investment in information technology is nothing new but has been an essential feature of capital markets since the days of Nathan Meyer Rothschild and his carrier pigeons. We also note that the rhetoric of new technologies, notably the internet, of equal access to information has not been fulfilled. In the days of HFT with its enormous technological infrastructure, public information is transformed into orders that are brought to the market extremely fast so that they resemble private information – at least with regard to other, slower, market participants. Forefronting the processes and strategies contained in the assemblages of HFT is essential in recognizing that the recreation of capital exchanges is not simply an exercise in efficiency but a calculated strategy. The human traders directing the efforts of HFT assemblages rely upon space-based strategies of information inequality to extract profits while simultaneously introduce new and unknown risks into the market. The earth’s geography and the speed of computer trading make information inequality an inherent characteristic of current market designs. As shown earlier, the practices within these assemblages are transmitted beyond the market into the larger society and assign non-market participants risk. They undermine assumptions of equal access to information necessary both for pricing and confidence in markets.

In reaction, policy-makers have suggested a number of changes to market design such as a cap on the frequency of trading, a minimum resting time for orders in the order book (Kay, 2012), and inserting small (up to a few microseconds) random delay to any order (Angel,

Figure 2. Nasdaq Composite Index, 22 August 2013.
2011). Perhaps, the most notable is the introduction of a financial transaction tax on each trade which would make the slim margins upon which high volume HFT strategies depend even thinner. Resistance from the human actants in HFT is strong (see Foresight, 2012) since much of the current activities of exchanges would be rendered moot and radically restructure the HFT assemblage. While all of these approaches would impact HFT trading strategies, none of them would solve the issue of information inequality inherent in HFT. For example, any financial transactions tax would need to be small in order not to unduly burden long-term investors. That means that it would still pay to be the fastest trader in the market during major news events (such as the release of key indicators outlined above) that move prices to a large extent.

Therefore, we call for a self-conscious and reflexive review of the assumptions and behaviors associated with capital market designs, for example, a radical rethinking of the rules of exchanges that are constructed for the benefit of certain traders. While there are any number of possible intervention points, we wish to highlight continuous trading (at least during operating hours) within exchanges. After all continuous trading is a design choice rather than a requirement, a point aptly made by an outage of Nasdaq on 22 August 2013. Due to technical problems, all trading on Nasdaq stopped abruptly on 12:14 pm and only resumed at 3:25 pm after the exchange conducted an auction to determine the prices of securities (see Figure 2).

Amazingly, it seems that this disruption in continuous trading had little effect on the market. One market participant quoted by Reuters said the reopening was “very orderly and liquidity is back to normal. It is almost like it did not happen” (Mikolajczak and Campos, 2013). The New York Times quotes an anonymous CEO of a Wall Street Firm: “We didn’t lose any money on the shutdown, but we also made very little money today” (Popper, 2013). “Public” and “equal” information would become more similar again when continuous trading is removed. In our view, this is exactly why it is an important exercise to explore the socially constructed norm of continuous trading and the current market design. After all, understanding the social basis for markets is the key first step towards affecting change in capital exchanges.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. While the market share of HFT varies between stock exchanges most studies find a significant share of HFT in total market turnover. Gomber et al. (2011, appendix I) reported figures of HFT’s share between 20 and 40 percent in European exchanges in 2010, with some estimates as high as 77 percent for continuous trading in the UK (see also Kay, 2012). The US reported similar figures – above 60 percent market share in 2011 – although with a subsequent drop to 50 percent in 2012.
2. Other factors in latency include serialization delay (time required to convert bytes of data from computer memory to a bit stream for transmission), data protocol handshakes and bit handling (checks on proper transmission and error handling), routing and switching (determining the pathway of bits) and buffering and congestion (delays caused by the number of bits reaching the capacity of the transmission medium) (o3bnetworks, n.d).

References


Domm P (2013) False rumor of explosion at White House causes stocks to briefly plunge; AP confirms its twitter feed was hacked, CNBC 23 April. Available at: www.cnbc.com/id/100646197 (accessed 14 August 2015).


