Government Office for Science

Foresight

Impact of special relativity on securities regulation

The Future of Computer Trading in Financial Markets - Foresight Driver Review – DR 15

Contents

Impact of special relativity on securities regulation

James J. Angel, Ph.D., CFA

McDonough School of Business
Georgetown University
Washington, DC 20057 United States
1 (202) 687- 3765
angelj@georgetown.edu

Draft 1.0
April 20, 2011

Comments welcome.

This review has been commissioned as part of the UK Government's Foresight Project, The Future of Computer Trading in Financial Markets. The views expressed do not represent the policy of any Government or organisation.

Introduction

Our markets move much faster than ever before. In a few short years, we have gone from a world in which humans traded face to face with humans to one in which computers trade with computers. Rather than responding in human-scale time, our markets now respond in computer-scale time. Traders fret about whether other traders have a millisecond advantage over them. Firms pay extra in order to have their computers co-located in the same data centers as stock exchanges, so that their trades are not delayed by the length of time it takes for an electronic signal to get from their office to the exchange.

What does this speed mean for our financial markets? In the early 20th century, physicists discovered that matter behaves differently at speeds approaching the speed of light than it does at lower speeds. Do our financial markets also behave differently as they approach trading at the speed of light? The goal of this exploration is to see whether the insights that physicists have gained from modern physics have relevance for how we regulate our financial markets. Are markets at the speed of light different?

This paper is organized as follows. The next section recounts a few of the well known but bizarre findings of modern physics. Space does not permit a complete picture of the elegant advances physicists have made in the last century, but a few of the more important findings are highlighted. In particular, information does not travel faster than the speed of light. Concepts of time need to be rethought. Locations of particles cannot be determined with certainty but instead become probabilistic smears.

Section II describes why speed matters to some participants in financial markets. Many traditional trading strategies such as market making or arbitrage involve a race between participants attempting to capture the same profitable trading opportunity. If one loses such a race by even one microsecond, one still loses. This motivates the participants to expend resources to speed up their trading systems so they win those races.

Section III discusses the regulatory implications of this race for speed. Can high-speed trading destabilize markets? Does it give some participants an unfair advantage? Our markets are now decentralized high- speed networks. Rules that attempt to force uniform prices at one moment in time over geographically decentralized places must take into account the fact that information does not move faster than light. At the same moment in time, two geographically separated participants may observe two different “best” prices. Attempts to enforce rules by time-stamping trades run into the problem that it becomes impractical with current technology to keep clocks synchronized to the nanosecond: regulators may never be able to establish that one event occurred before or after another event.

I A brief recap of special relativity and modern physics

Prior to the 19th and 20th centuries, the prevailing scientific paradigm held that the universe mechanically obeyed Newton’s laws of motions. In a Newtonian world, velocities add up. For example, if one is travelling on a train and shoots a bullet in the same direction as the train is travelling, the speed of the bullet to an observer standing still outside the train would be the sum of the speed of the train and the speed of the bullet.

The speed of light is a speed limit on the flow of information.

Discoveries in the late 19th and early 20th century challenged this view. The universe does not behave in a strictly Newtonian manner. For example, experiments showed that the speed of light is a constant, regardless of whether the light source is moving towards the observer or away from the observer. Suppose one person travelling on a high-speed train shines a flashlight in the same direction as the train is travelling. A stationary observer outside the train would see the light beam moving at the same speed as observed by the person on the train, despite the fact that the light source itself was moving forward. The outside observer does NOT observe light travelling at the speed observed by the inside observer plus the velocity of the train. Similarly, velocities do not add even for non-light objects. If the train was travelling at 2/3rds of the speed of light and it fired a bullet forward at 2/3rds of the speed of light, the speed of the bullet would not be 4/3rds of the speed of light to the stationary observer.

This invariance of the speed of light has some important implications. Nothing moves faster than light in a vacuum.[1] Under special relativity, as a particle accelerates toward the speed of light, it gets heavier. It would take a theoretically infinite amount of energy to accelerate it all the way to the speed of light. The speed of light thus becomes a speed limit on the transmission of information as well. This has important implications for financial markets regulation.

Time does strange things at relativistic speeds.

This invariance of the speed of light has implications for the concept of time. One of the interesting findings of special relativity is that as an object moves faster time slows down. [2] This leads to the famous "twins paradox." Suppose there are two twins born on the same day. One gets on a rocket ship that travels at close the speed of light while the other stays home on planet Earth. When the rocket traveler returns to Earth, he will actually be younger than the twin who was left behind on Earth. However, special relativity is not just of use to space travelers. GPS units on earth need to adjust for fact that time flows at different rates at different altitudes.

Closely related is the relativity of simultaneity. Because information can only travel at the speed of light, two events that appear to happen at the same time to one observer may appear to happen at different times to a second observer in another reference frame. Special relativity also demonstrates the equivalence of mass and energy, given by Einstein's famous equation $e = mc^2$.

Quantum mechanics highlights the problems of measurement at a small scale

Quantum mechanics is another realm of modern physics where interesting things happen that are very different from what we commonly observe in the macro-world. Some particles can inhabit discrete quantized states and not be in between. Furthermore, the Heisenberg uncertainty principle shows that there is a limit to the precision with which the location and momentum of a particle can be determined. Particles don't necessarily exist in any one place – they are really a probabilistic smear. A particle could be in different locations at the same time, just with different probabilities.

---

[1] This is somewhat of an oversimplification since in quantum mechanics some virtual particles can move faster than light but do not transmit information.

[2] This has been demonstrated experimentally by taking very accurate atomic clocks on high-speed jet rides in different directions. See J.C. Hafele and R. E. Keating, Science 177, 166 (1972)

Not only is the location of a particle a probability function, so are other states of a quantum system. A quantum system may exist as a “superposition” of various states, and only when the system is observed does the system become one state or another. This is best described by the famous “Schrodinger’s cat” thought example in which a cat is sealed in a box and its life is dependent on the state of a quantum mechanical system which may exist as a superposition of different quantum mechanical states. The cat is both alive AND dead until someone opens the box to observe the state of the system. It is only then that the state of the system – and the cat’s life -- is determined.

Light is both a wave and a particle.

Another important concept in physics is the duality of electromagnetic radiation. Light has the properties of both particles and waves. Sometimes light acts like a collection of discrete photons, but it also behaves like a wave.

Not even light can escape from a black hole.

Gravity imposes its tug on light as well as regular matter. Under everyday conditions, the impact of gravity on light is not noticeable. However, some objects are so massive and so dense that their gravity is strong enough to prevent light from escaping. These are known as black holes. Astronomers now believe that there are black holes at the center of many, if not most, galaxies.

Nonlinear systems can be chaotic.

Many physical systems are highly nonlinear. A small change in one input variable at one time can lead to a miniscule change in the system at one time or a disproportionate change at other times. Chaos theory has shown that some physical systems display such a sensitive dependence on initial conditions that, even though they are mathematically predictable, they are practically unpredictable. Various weather models display this kind of sensitivity to the initial conditions, leading to the famous “butterfly” effect in which the flapping of the wings of a butterfly can theoretically lead to -- or prevent – a subsequent tornado.

The important lesson that we take from physics is that assumptions based on observation of everyday phenomenon need to be modified under extreme conditions of high velocity or at very small scales. It would be useful to have a realistic model of financial markets that starts from the first principles of modern physics. Such a model could be used to explore the impact of trading at higher speeds on market dynamics and stability. However, such a model does not yet exist. In the meantime, we can use the insights we have gleaned from physics to help us think more broadly about the challenges involving high- speed trading.

II Why speed matters

In order to understand the regulatory implications of high-speed trading, it is important to understand what high-speed traders are doing. This section describes many of the strategies involved in high-speed trading. Market participants engage in a wide variety of trading strategies, some of which involve reacting quickly to market conditions and trading rapidly. These so-called “high frequency” traders are often implementing some very old and fairly low-tech strategies. Examples of high-frequency strategies include:

Arbitrage

Arbitrageurs attempt to profit from the relative mispricing of two related securities. For example, Royal Dutch Shell has two classes of common stock outstanding, class A and class B. The two classes have identical rights except that the dividends for the A shares are from Dutch sources and the dividends for the B shares are from UK sources. When the price of one class deviates from the other, an arbitrageur can profit by buying the cheaper share and selling the more expensive share.[3] This activity tends to push prices back towards their proper alignment. After the prices of the two shares move back into alignment, the arbitrageur closes out the position to reap the profit.

Similarly, arbitrage activity keeps the prices of a wide variety of derivative products properly tied to related instruments. For example, arbitrageurs keep the prices of exchange traded funds (ETFs) in the proper alignment with the prices of the stocks that are held by the ETF. This means that retail investors can safely buy and sell ETFs as this arbitrage activity keeps the prices of the ETF where they should be relative to the underlying securities. Arbitrageurs similarly keep the prices of options and futures in the proper alignment with the underlying assets.

Arbitrage trading is a rather simple and low-tech strategy: Program the computers to watch the prices of the two related instruments and when the prices diverge, buy low and sell high. Because of this simplicity and the ease with which anyone can begin trading, there is much competition. When an arbitrage situation arises, there is a race to snap up a profitable trading opportunity before it disappears. An arbitrageur who comes in second in the race for a profitable trade still loses, whether they lose by one second or by one nanosecond. For this reason, such traders invest heavily to make sure that their trading systems respond as fast as possible to market conditions. Exchanges have sped up their response times and major exchanges now have "latencies" of around one millisecond (one thousandth of a second) or less. Exchanges and practitioners now routinely time stamp their messages to the millisecond. Some firms pay extra to co-locate their servers in stock exchange data centers so that their orders can get to the exchange's matching engine sooner. In one millisecond, light travels about 300 km, roughly the distance between London and Paris. (In real computer networks, the signal travels a little slower because of the delay in going through a solid and because of switching delays in the network.) In one microsecond (one millionth of a second), light travels about 300 meters. In one nanosecond (one billionth of a second), light travels about 30 cm. In a very close race between two computerized traders, that nanosecond just might make the difference between catching a profitable trade and missing it.

Market making

Another time-honored and low-tech strategy is market making. In real life, customer buy orders and sell orders do not always arrive at the exchange at the same time. Just as used car dealers provide immediacy to people who want to buy or sell a used car quickly, market makers apply the same used-car dealer principle to trading securities. They stand willing to buy or sell a security as an accommodation to other market participants – for a price. Market makers simultaneously enter bids to purchase securities and offers to sell them at higher price. Just as used car dealers do not want to actually keep the cars they buy and sell, market makers do not want to hold an inventory position very long. Holding inventory ties up capital and it involves risk. Market makers profit from the difference between the low bid price at which they buy and the higher offer price at which they sell. Their willingness to trade makes it possible for investors to buy or sell whenever they want without having to wait a long time to find the contraparty.

---

[3] If the arbitrageur does not already own the more expensive share, he or she can still sell it short by borrowing the shares from another shareholder. The arbitrageur would later buy back the shares in the market to repay the stock loan.

Market making is also a very competitive field. Since stock computers usually fill orders on a first-come, first-served basis, it is important to be able to enter orders quickly. Furthermore, when it is apparent that the market price is changing, it is important to be able to cancel orders quickly to avoid trading at the wrong price just before the market moves. This is one of the reasons why some high frequency traders cancel many of the orders that they place. Again, speed is important, both for submitting orders and for cancelling them.

Pattern recognition

Pattern recognizers attempt to spot patterns in stock prices and trade accordingly. This is the modern form of the old tape watchers and technical analysts who used to watch the ticker tape in brokerage offices. They use the information on recent trades to help them forecast which way prices will go. Again, speed is of the essence as one's competitors may spot the same trend at the same time. Some large institutional "buy side" traders call such pattern recognizers parasitic traders because one of the patterns they recognize is when a large institution is buying a large block in pieces. The pattern recognizers make use of the price and volume patterns they see from the pieces of the block to guess that a block is being accumulated in bits. The pattern recognizers infer that the institution knows something and will continue to move the price up, so they buy alongside the institution. This pushes up the price the institution has to pay.

News reaction

News processors electronically process news feeds and trade on the news. Once again, this is an old strategy that uses computers to do the job more quickly than in the old days. Speed is critical to this type of trading as other investors, both human and machine, are processing the same news and will soon trade on it.

Manipulation

Manipulators attempt to move prices away from their fundamental values in the hopes of making a short-term profit. For example, when manipulators sense that a stock is vulnerable, they may place a large order to sell short. The sell order may trigger selling by other investors (through momentum trading by other investors, pre-existing stop orders to sell when a stock drops below a given price, or margin related selling when investors have borrowed to finance the purchase) and result in a self-fulfilling (and profitable) decline in the price.

III Regulatory concerns about high-speed markets

Modern physics has made it clear that matter and time behave differently as speed approaches the velocity of light. We can use the insights of physics to see what pitfalls may occur in the understanding and regulation of high-speed markets.

Financial market regulation has multiple – and often conflicting -- objectives.[4] These include consumer protection, fairness, resource allocation, economic efficiency, capital formation, soundness of financial institutions, and economic stability.

---

[4] For example, see Conflicting Objectives of Financial Regulation, by Richard Aspinwall; Challenge, Vol. 36, 1993

Consumer protection: Does trading at the speed of light have implications for consumer protection?

Public policy decisions in the United States and Europe have resulted in competitive equity markets. As the old monopoly markets privatized, public policies such as Regulation NMS in the United States and MiFID I&II in Europe adopted pro-competition policies to prevent the economic problems resulting from for-profit monopolies. New exchanges and multilateral trading platforms have arisen, resulting in a much more competitive trading environment. Brokerage firms also provide competition by trading directly with their customers rather than sending orders to other exchanges. This has resulted in equity market networks that consist of geographically diverse trading nodes. While this combination of competition and technology has generally resulted in improvements in measures of market quality, there are concerns.[5]

Best execution and trade through rules

In a world of competing trading venues, some clients have difficulty determining whether their brokers have done a good job executing their trades. Although brokers face strong commercial pressure to deliver good service to their customers, they may also benefit financially in ways that harm their clients. For example, a broker may route an order to a trading venue that has an inferior price but that charges the broker lower fees or even pays a rebate to the broker. Not all customers have the capacity to monitor closely execution quality. Regulators seek to protect investors by requiring that brokers seek "best execution."[6] While this principle of best execution is somewhat vague, it can be interpreted as getting the best price available at a given moment. A similar rule in the United States is the "trade through" rule of Regulation NMS that requires exchanges to not "trade through" the quotes of another exchange. For example, suppose one exchange is offering to sell stock at 11 and another exchange is offering to sell stock at 10. If the first exchange receives a buy order, it is supposed to route the order to the exchange with a better price rather than filling the order itself at an inferior price. (Note that orders and trades, like photons of light, are discrete quantum phenomena, even though in common parlance we may speak of a "wave" of sell orders.)

Enforcing such rules in a high-speed world runs into two problems. The first problem stems from the fact that information does not travel faster than light. If the two exchanges above are geographically separated, the information that one exchange has a better price may not have reached the other exchange before the order arrives.[7] In the previous example, suppose the first exchange receives an order to sell stock at 9 and updates its quote, which is broadcast to the market. At the instant the new quote is broadcast but before the signal reaches the second exchange, observers at the two separate exchanges will observe two different "best" prices. An observer at the first exchange sees the best price as 9 and an observer at the second exchange sees the best price as 10.

---

[5] See Angel, James J., Harris, Lawrence and Spatt, Chester S., Equity Trading in the 21st Century (February 23, 2010). Marshall School of Business Working Paper No. FBE 09-10. Available at SSRN: http://ssrn.com/abstract=1584026.

[6] For example, see the MiFID best execution requirement in MiFID Article 21, discussed in http://www.fsa.gov.uk/pubs/discussion/dp06_03.pdf. See also Macy and O'Hara, "Law and Economics of Best Execution", *Journal of Financial Intermediation* 6, July 1997 188-223, for a discussion of the US situation.

[7] The implications of such light propagation delays are explored in Wissner-Gross, A.D, and C.E. Freer, Relativistic Statistical Arbitrage, Physical *Review E* 82, 2010, DOI: 10.1103/PhysRevE.82.056104.

Alas, the second exchange receives the order to buy before the updated information that there is a better offer at the other exchange arrives. The second exchange then fills the order at its price of 10. From its perspective, it has given the customer best execution. However, an observer at the first exchange would say that the second exchange has traded through the first exchange and should be punished.

How would a regulator determine whether the exchange had broken a “trade-through” rule or not? One regulatory solution is to require market participants to time stamp all market events. However, there are limits to the practical precision which can be obtained at reasonable cost by market participants. Syncing to a standard time source such as an atomic clock sounds simple in theory, but there are limits to the practical accuracy of what can be implemented. Even with access to a signal from a standard time source, there are still random propagation delays within the processing hardware that can limit the precision of the resulting time stamps. Thus, a regulator attempting to reconstruct the sequence of events and determine whether one exchange filled the order before or after the other exchange displayed a better price may never be able to accurately determine what happened.

In a similar vein, some have called for time priority to be preserved across different trading venues: they believe that the first to post a quote at a particular price should get the first opportunity to be filled. This is simple to enforce within a single exchange’s computer. However, enforcing such a rule across disparate platforms in different geographic locations runs into the same problem as a trade through rule. Because information can only travel at the speed of life, the current state of the market will appear differently in different geographic locations. A regulator attempting to determine whether a trading platform is obeying the rule has a Heisenbergian uncertainty about the where the market is at a particular point in time. This makes monitoring and enforcement of such a rule problematic.

Fairness: Is trading at the speed of light fair?

Some critics claim that it is unfair that some firms are allowed to “co-locate” their computers in stock exchange data centers. This proximity to the exchange matching engines allows them to receive market data faster and to submit orders faster than other participants located farther away. It is true in physics that particles moving faster have more energy. Similarly, traders who can trade faster have more power. However, this power is available to anyone who wishes to pay for it. As long as such services are available on fair and non-discriminatory terms to all market participants, there is nothing especially unfair about the practice of co-location. For centuries brokerage firms have located offices as close as possible to the exchange so that they could get their orders into the exchange faster. The only difference is that the orders are now submitted via electrons and not runners.

Is it fair that some participants have the resources to spend on co-location that others don’t? It is no more unfair than the fact the some investors are endowed with more resources to spend on fundamental research, or better brains for finding good investments.
If the practice of co-location were banned, the co-locators would merely move to another location as close as possible to the exchange data center. Rather than the exchange getting the revenue, the landlord of the other location would collect more rent.

Resource allocation: Does trading at the speed of light misallocate resources?

A misallocation of resources would occur if either the trading mechanism itself consumed excessive resources or resulted in bad prices that led to a serious misallocation of resources. In general, trading with computers is much cheaper than trading with humans, so transaction costs have fallen steadily in recent years as a result of the automation of the markets. The intense competition between electronic participants has resulted in dramatic drops in trading costs in recent years.

One could argue that the resources expended by the high frequency traders to speed up their systems are socially wasteful. Wouldn't the markets be just as efficient if their response times were a few milliseconds slower? Perhaps. However, attempts to slow down the race are likely to be unsuccessful. Even if the exchanges themselves slowed down, the participants still have the same strong financial incentives to beat their competitors by a nanosecond. Attempts to tax high frequency trading out of existence would likely eliminate the beneficial trading along with non-beneficial trading, driving up transactions costs for all investors.
Whether electronic trading creates bad prices that misallocate resources is another matter. Financial markets have always been prone to bubbles, and it is not clear that high-speed trading would make such bubbles any worse.

Economic efficiency: Does trading at the speed of light harm economic efficiency?

Some high-frequency trading strategies such as arbitrage and market making clearly help the market to operate more efficiently. News reaction strategies help the market to incorporate information faster. On the other hand, manipulators can also use computers to implement manipulative strategies.

So-called dark pools have arisen in recent years that limit display of their information. Unlike exchanges that disseminate their limit order books, dark pools typically do not display such information. It is an old adage in the markets that "liquidity attracts liquidity." Dark pools have grown dramatically in recent years. Could a dark pool attract so much liquidity and order flow that it becomes a black hole that takes over the market?

Soundness of financial institutions: Does trading at the speed of light threaten the soundness of financial institutions?

One of the lessons of quantum physics is that occasionally extreme events can and do occur. Particles can "quantum tunnel" through a solid barrier even when they do not have enough energy to get through it by classical calculations. There is a nonzero probability that a particle such as an electron can get through a barrier no matter how high the barrier.

In our highly complex and nonlinear market network, freak events will occur. These can occur from programming glitches in which seemingly innocuous minor changes lead to a major catastrophe. For example, suppose that a brokerage firm rolls out a new software patch to meet a deadline to comply with a mandated rule change. The change passes through the usual quality assurance testing. However, an obscure glitch could suddenly result in magnifying the orders of some customers by 100, turning 100 share sell orders into 10,000 share orders. The excess sell orders ignite a cascade of selling activity and triggers a crash. By the time the glitch is discovered and the system turned off, the brokerage firm fails.

Such catastrophic freak failures would have been caught earlier in the much more expensive days when every order passed through human hands. Regulators need to be vigilant about such risks and have contingency plans in place for containing the damage when such events happen. Murphy's Law guarantees that they will happen.

Economic stability: Does trading at the speed of light threaten financial stability?

Most high frequency trading strategies tend to be stabilizing to the financial markets. Market making provides buying power when others want to sell and sells when others want to buy. This negative feedback tends to stabilize prices. Arbitrage keeps the prices of related instruments in the proper alignment. However, other activities can destabilize the market.

Momentum traders tend to buy when stocks are going up and sell when they are going down, which can accelerate volatile price moves.

Under the right combination of events, the nonlinear market network can become unstable, as demonstrated by the US experience of the "Flash Crash" on May 6, 2010.[8] On a day of heavy trading volume and generally falling prices, a large investment firm placed a large sell order on futures contracts tied to the S&P 500 and the price of the contract plummeted. There were reports of data problems at several firms. Several high frequency trading firms withdrew from trading, citing "data integrity" problems: given their need to respond quickly to market conditions, their reliance on accurate and timely market data, and their generally thin profit margins, it was only prudent for them to withdraw from the market when they suspected problems with their data feeds. The lack of market making and arbitrage activity, while other investors continued to trade, caused highly erratic cash equity prices. The market quickly recovered, but over 20,000 "clearly erroneous" trades were cancelled.

Even under normal circumstances, there are concerns that high-speed trading may exacerbate market volatility. At various times the various computer programs may interact in such a way as to cause a temporary shortage of liquidity which would increase volatility.

Capital formation: Does trading at the speed of light harm capital formation?
By lowering transactions costs, electronic trading should in theory actually aid capital formation. The ability to sell in a secondary market with lower trading costs and higher liquidity should make such assets more attractive to investors in the primary market, thus enhancing capital formation. On the other hand, if high-speed trading increases market volatility, then it could have a deleterious on capital formation. The net effect of high-speed trading on capital formation is thus an empirical question.

[8] See the SEC/CFTC joint reports. http://www.sec.gov/sec-cftc-prelimreport.pdf and http://www.sec.gov/news/studies/2010/marketevents-report.pdf.

Conclusions and Summary

The speed of our financial markets has accelerated in recent years as we have moved from human- intermediated markets to machine-intermediated markets. Furthermore, our markets are now geographically diverse complex networks of competing trading platforms. This shift brings up several issues regarding the regulation of our high-speed markets. Modern physics has taught us that objects moving at high speed behave differently from objects moving at low speed. Does the same hold true for our financial markets?

The speed of light provides an upper limit on the speed with which information can flow. This has large implications for any rules that are based on prices at a given time. Two observers in two different locations can simultaneously observe two different “best” prices. Similarly, the inherent limits to the precision of time measurement mean that there will be an irreducible uncertainty as to when particular events occurred. Again, this will have serious implications for implementation and enforcement of rules seeking to ensure that a customer gets the best price available at a given time.

The highly nonlinear nature of the financial market network is also cause for concern. Under normal conditions many financial markets participants provide negative feedback that tends to stabilize the markets as they buy low and sell high. Others, such as momentum investors add positive feedback to the market. Conditions can occur that lead to temporarily destabilizing positive feedback.

Foresight
1 Victoria Street
London SW1H 0ET
www.bis.gov.uk/foresight
URN 11/1242