

The Flash Boys of Wall Street

The Flash Boys of Wall Street:

How High-Frequency Programmed Trading, Black Pools,
Network Latency and Front-Running Rigged the Financial
Markets in the 2010's

By Bob Primak

For The Lexington Computers and Technology Group (LCTG)

Wed., June 22, 2022, 10:00 AM EDT

The Flash Boys of Wall Street

Based on the New York Times best-selling non-fiction book by
Michael Lewis

Featuring research and testimony from Brad Katsuyama,
and technical expertise from Ronan Ryan and others.

<https://www.amazon.com/Flash-Boys-Wall-Street-Revolt/dp/0393351599>

The Flash Boys of Wall Street

What is high-frequency trading?

<https://www.cnbc.com/2014/03/31/what-is-high-frequency-trading.html>

High-frequency trading is the use of a highly sophisticated computerized trading offering an edge to those who use it. Computer algorithms take a look at a variety of factors including economic reports and news headlines to make split-second determinations.

The Flash Boys of Wall Street

Critics of high-frequency trading, including Mark Cuban and author Michael Lewis, have recently brought this risky, powerful trading into the spotlight. In May 2010, when the Dow dropped more than 1,000 points (and regained them) in less than 20 minutes, this “flash crash” was largely attributed to high-frequency trading.

The Flash Boys of Wall Street

High-frequency trading was not regulated in 1997 when it first began to become popular among Wall Street brokerage firms. Since then, the SEC (Securities Exchange Commission) and the CFTC (Commodities Futures Trading Commission) have introduced some regulations, including limit-up, limit-down, daily and hourly limits and other safeguards, in light of the “black bag mortgage” crisis (which precipitated the decline of 2008-2010), the “flash-crash” of 2010 and other events in which HFT was said to play major roles.

The Flash Boys of Wall Street

HFT is not the same thing as “smart contracts” in a blockchain, which are coming under tremendous scrutiny as NFT and cryptocurrency (and stablecoin) ecosystems are crashing down now. The term “programmed trading” does not really describe HFT, as unlike the blockchains and their smart contracts, there are points where human intervention and regulatory rules and actions can interrupt unwise or downright catastrophic gyrations in the valuation of market assets.

The Flash Boys of Wall Street

See my previous talk on cryptocurrencies and blockchains

Cryptocurrency and Validated Transactions By Bob Primak For
The Chicago Computer Society South Suburban Computer Club
May 17, 2022.

Slides 14-21, and related references.

(I'll post this document (PDF) to Chat.)

The point is, this is a different and older form of programmed trading.

The Flash Boys of Wall Street

Today, much of this activity is regulated by the SEC and CFTC. But regulation and human intervention were not the rule in the Wall Street of 2007-2013, when this story took place.

The Flash Boys of Wall Street

NY Times Journalist Michael Lewis said that the problems in the market structure “screw-up” were “unintended consequences” from regulations introduced in 2007.

“It’s an accidental conspiracy,” Lewis said.

“I think the high-frequency traders are like exploiting a system that’s got these glitches in it, and they found loopholes to jump through. It’s sort of like blaming the lion for eating the antelope that they do what they do,” he added.

<https://www.cnbc.com/2014/04/01/michael-lewis-rigged-markets-an-accidental-conspiracy.html>

The Flash Boys of Wall Street

How does High Frequency Trading work?

“I’ve said several times that I do not believe the markets are ‘rigged’ or that high-frequency trading is legalized front running. However, I have said that it would not surprise me that there are high-frequency traders who are engaging in abusive or manipulative behavior.”

-- Bob Pisani (CNBC Commentator) -- Lost in all the ‘Flash Boys’ talk

<https://www.cnbc.com/2015/03/23/pisani-lost-in-all-the-flash-boys-talk.html>

The Flash Boys of Wall Street

(Pisani continues)

The real focus should be on the strange relationships that have developed between brokers, dark pools, exchanges, and high-frequency traders that has led to the markets being excessively complex, fragmented, and confusing.

I'm talking about developments like paying traders to trade on an exchange, which has led to the proliferation of strange order types that route stocks in confusing directions.

The Flash Boys of Wall Street

Including:

- dark pools that do not adequately disclose who is participating in their pools
- the obsession with getting as close to possible to exchange servers in order to have a microsecond speed advantage in trading
- the development of trading strategies designed to exploit sub-second strategies, like the use of pattern recognition software to determine the location of large buyers or sellers

The Flash Boys of Wall Street

This all raises legitimate questions about whether the system is needlessly complex and, if it is, how to reduce the complexity or reduce the number of ways the complexity can be exploited.

Lost in all the 'Flash Boys' talk

<https://www.cnbc.com/2015/03/23/pisani-lost-in-all-the-flash-boys-talk.html>

The Flash Boys of Wall Street

The Network Latency Arms Race of the 2010s

In 2010, a company named Spread Networks spent \$300m to dig a fiber-optic line between Chicago and New York. The sole purpose of this very expensive cable? Reducing latency – the time it takes to receive and respond to new information – for trading firms. Less than two years after its completion, however, the wire was rendered obsolete by more sophisticated technology: a communication network transmitting microwave beams through the air.

The Flash Boys of Wall Street

This one-upmanship is just one symptom of the latency arms race that's been perpetuated by high-frequency trading (HFT) – a phenomenon that's been leading to problems long before Michael Lewis dropped his latest bestseller this week. HFT is characterized by traders operating at the microsecond level, using automated computer algorithms to submit large numbers of orders to buy or sell.

Some firms use co-location, placing their computers as close an exchange's servers as possible. Others rely on dedicated communication lines from companies like Spread Networks. Still more firms pay for direct feeds from exchanges – and they pay a lot.

The Flash Boys of Wall Street

Lewis contends that the US stock market "really was rigged"—indeed, that it still is, with high-frequency traders exploiting their latency advantages at the expense of Average Joe investors. Well, the US stock market may indeed be rigged, but it's not just about the high-speed trading. The stock market is rigged because of how current stock markets work, and my research has found that it's rigged against the traders themselves.

The Flash Boys of Wall Street

Virtually all modern financial markets match orders continuously – that is, as orders arrive to the exchange. Continuous-time matching is essentially a winner-takes-all race. A high-frequency trader who receives and acts on new information faster than others can readily pick off orders sitting on exchanges – over 40 venues are competing for the same orders – before others can react. So being faster by as little as one microsecond is enough to grab all the profit.

This is how the "flash boys" win.

The Flash Boys of Wall Street

HFTs employing latency arbitrage examine current market information to predict immediate price movements, essentially computing the best prices available before the exchange has even had a chance to update its price quote.

The Flash Boys of Wall Street

What was found in a 2013 University of Michigan study may be even scarier. (The article, like Lewis's book, was published in 2014.)

It's not simply a matter of the HFT crowd taking profits away from regular investors. Predatory strategies like latency arbitrage have the potential to reduce trading gains for all market participants – high-frequency traders and Average Joes alike.

The Flash Boys of Wall Street

In Flash Boys, Lewis tells the story of IEX, a stock exchange designed specifically to protect investors from predatory HFT strategies like latency arbitrage. The founders of IEX intentionally introduced a 350-microsecond delay to ensure that high-speed traders could not act on updated information from the exchange before regular investors. These are the heroes, we are told.

The Flash Boys of Wall Street

But the dominance of sophisticated computer-based trading algorithms is due to the nature of continuous-time trading, and imposing a fixed delay may be at best a stopgap measure. The alternative is to match orders to trade at fixed, regular intervals, like every second.

In such a "frequent call" market, everyone has until the next time the exchange clears, or matches orders, to respond to new information. Orders are processed at the same time, rather than as they arrive, so the fastest trader no longer takes everything. The interval length can be selected to be small enough to effectively track information in the markets, yet imperceptible to regular investors.

The Flash Boys of Wall Street

Similar to the idea behind IEX, the frequent call market slows down trading by introducing a delay. And matching at regular intervals provides a notable additional benefit: our study showed that frequent call markets offer a significant improvement in trading gains.

As Michael Lewis explains so well, exploitative Wall Street behavior has run amuck, eroding trust and investor confidence. But we need to understand the vicious cycle happening at the exchanges – not just the HFT arms race – to truly escape the rigged game.

The Flash Boys of Wall Street

Michael Lewis says the market's rigged. But his 'Flash Boys' rigged themselves

<https://www.theguardian.com/commentisfree/2014/apr/04/michael-lewis-market-rigged-flash-boys-high-speed-trading>

The Problem's Roots: Latency Arbitrage

What Is Arbitrage?

The Flash Boys of Wall Street

Arbitrage is an investment strategy in which an investor simultaneously buys and sells an asset in different markets to take advantage of a price difference and generate a profit. While price differences are typically small and short-lived, the returns can be impressive when multiplied by a large volume. Arbitrage is commonly leveraged by hedge funds and other sophisticated investors.

There are several types of arbitrage, including pure arbitrage, merger arbitrage, and convertible arbitrage.

The Flash Boys of Wall Street

In other words, there is an asking price and there is an offering price. Arbitrage is the process of reconciling the two price levels to facilitate a sales transaction or contract. A fee is paid to the arbitrageur for this service.

What Is Arbitrage? 3 Strategies To Know

<https://online.hbs.edu/blog/post/what-is-arbitrage>

Arbitrage Funds – Meaning, Basics, Things to Consider & More

<https://cleartax.in/s/arbitrage-funds>

The Flash Boys of Wall Street

Arbitrage differs from blockchain “smart contracts” in many ways. I outlined some of the differences earlier.

There are Arbitrage Funds – Mutual Funds which are focused on arbitrage, and you can invest in them (though I would not recommend doing so).

Effects of Latency Arbitrage

Traders and hedge funds who use high-speed methods to gain an advantage in the stock market impose a “tax” on other investors, according to a study released in 2020, costing as much as \$5 billion per year across global exchanges at that time.

The Flash Boys of Wall Street

The Financial Conduct Authority (FCA), a regulatory arm of the United Kingdom, found that the trading practice, known as “latency arbitrage,” causes the overall volume of trading on global stock markets to decrease.

Latency arbitrage is one of the ways high-frequency traders profit to the detriment of slower trading investors. It involves arbitraging prices gleaned with a low latency -- in fractions of a second -- from certain exchanges. Better prices are snatched up by high frequency traders before regular investors.

The FCA found the average race between firms lasted 79 microseconds (79 millionths of a second), faster than the blink of an eye, with only the quickest to execute its trade gaining any benefit.

The Flash Boys of Wall Street

This is where network latency effects can enter into the problem.

While latency arbitrage trades on public information, the study found the negative outcome of such high speed trading is that it increases the cost for investors to buy and sell shares.

So technically, this is not insider trading or the illegal practice of front-running. But it is considered very unethical, and it has led to the development of “off-exchange” trades in so-called “dark pools”.

The Flash Boys of Wall Street

High-speed traders cost regular investors almost \$5 billion a year, study says (2020)

<https://www.cnbc.com/2020/01/27/latency-arbitrage-trading-costs-investors-5-billion-a-year-study.html>

One of the key players in our story is BATS, a non-traditional exchange located in northern New Jersey. (This exchange is today not what it was at the time.)

What Was BATS Global Markets?

The Flash Boys of Wall Street

Bats Global Markets was a U.S.-based exchange that listed several different types of investments, including equities, options, and foreign exchange. It was founded in 2005 and was acquired by Cboe Options Exchange (Cboe) in 2017. Prior to being acquired, Bats Global Market was one of the largest U.S. exchanges and well known for its services to broker-dealers as well as retail and institutional investors.

As an exchange, Bats grew into the main competitor to the New York Stock Exchange (NYSE) and Nasdaq, both of which handled a greater amount of equities when ranked by market capitalization. In 2016, Bats had become the second-largest U.S. equity exchange by market share and was the largest exchange-traded fund (ETF) exchange.

The Flash Boys of Wall Street

In 2013, the company indicated that a technical error led to hundreds of thousands of trades executing at prices lower than the best bid and offer, which also affected investors who were selling shares short. The error affected trades going back four years.

Cboe, the owner of the Cboe Options Exchange and Cboe Futures Exchange (CFE), made an offer to acquire Bats Global Markets in 2017. The acquisition allowed Cboe to expand into Europe and increase its offerings to include foreign exchange and ETFs. Cboe now operates four U.S. options markets, Cboe Futures Exchange, a European equities market, four U.S. equities markets, and a foreign exchange market.

Three of the exchanges that Cboe operated prior to acquiring Bats migrated to the Bats trading platform.

The Flash Boys of Wall Street

Better Alternative Trading System (BATS)

<https://www.investopedia.com/terms/b/better-alternative-trading-system-bats.asp>

More key terms – Front Running and Insider Trading

Both front-running and insider trading impact mutual fund unit holders, as well as stock market investors.

Front-running is a bit different from insider trading, although in both the cases, the perpetrators aim to make money on the stock market by trading in a company's shares. In front-running, a dealer within an institutional money manager like a mutual fund or even a share broker takes advantage of his knowledge of the orders that the mutual fund has lined up for the day and tries to profit from them.

The Flash Boys of Wall Street

How is a front-running fraud orchestrated?

Once the fund manager decides what he wants to buy or sell, he informs his dealer, whose responsibility it is then to execute the trades on behalf of the fund house. Every fund house has an isolated area called the dealing room, in which these trades are executed, all day long. Only dealers and fund managers are allowed entry here.

And here's where the crime can take place.

The Flash Boys of Wall Street

If the dealer wants to profit, he enters the market minutes before he punches in the order of the fund house. Mutual funds usually place large orders in the stock market. Such orders can sharply move the price of a stock.

Here, the dealer buys or sells the stock minutes before a mutual fund places its trades, buying or selling the stock.

The idea is to profit from the big investor's moves, either by buying or selling shares.

The Flash Boys of Wall Street

What is insider trading?

Insider trading, on the other hand, is when a company insider, an official, employee or a senior executive, takes advantage of unpublished price-sensitive information (UPSI) to trade in the company's stock and make profits from such transactions.

As the name suggests, insider trading is done by a company's employee aiming to profit from dealing in the company's shares. Front running can be done in just about any stocks or sectors by unrelated people, who have knowledge of how some large investors plan to trade in the markets.

The Flash Boys of Wall Street

“Work from home has made it difficult to keep controls”

Axis Mutual Fund under scanner: What is the difference between front-running and insider trading?

<https://www.moneycontrol.com/news/business/personal-finance/axis-mutual-fund-irregularities-what-is-the-difference-between-front-running-and-insider-trading-8477401.html>

What is Tailgating?

Tailgating is when brokers or financial advisors profit by placing orders on their own account using information provided by customers for their trades.

The Flash Boys of Wall Street

It is not illegal but is considered highly unethical.

Even though tailgating is not illegal, the SEC can take action against firms that make profits using information provided to them by customers.

Tailgating

<https://www.investopedia.com/terms/t/tailgating.asp>

The Flash Boys of Wall Street

https://personalpages.manchester.ac.uk/staff/m.dodge/cybergeography/atlas/more_isp_maps.html

The Wolf Hunters of Wall Street [*Paywall*]

https://www.nytimes.com/2014/04/06/magazine/flash-boys-michael-lewis.html?_r=0



The Flash Boys of Wall Street

How Point of Presence (PoP) affects latency, and what colocation means

What is a Point of Presence (PoP)?

<https://blog.stackpath.com/point-of-presence/>

Points of presence vs. edge locations

A point of presence can be as simple as “a single server mounted in someone else’s cabinet.” Edge locations, on the other hand, are points of presence with full deployments of advanced infrastructure—not just a single server.

The Flash Boys of Wall Street

How Points of Presence Work

Points of presence work very much like Internet exchange points, only on a smaller scale. The equipment located in a point of presence can be broken down into five categories:

Base stations: A central point of connection with an access point and bandwidth management to distribute connection speeds

Client equipment: Used by customers to connect to the central point of connection

Network switches: Provides the link to the “last mile” and is used for distribution

Routers– Provides multiple routes to the network

Firewall– Protects from internal and external threats

The Flash Boys of Wall Street

Points of presence include routers, switches, servers, and other devices necessary for traffic to cross over networks.

A point of presence can be as simple as “a single server mounted in someone else’s cabinet.”

Edge locations, on the other hand, are full deployments of advanced infrastructure next to IXPs.

The Flash Boys of Wall Street

Consolidation! The tale of five boroughs and one big city

<https://www.boweryboyshistory.com/2013/04/consolidation-tale-of-five-boroughs-and.html>

This is a reference for a map and brief history of New York City's Five Boroughs.

(next slide)

The Flash Boys of Wall Street



The Flash Boys of Wall Street

The Greater New York Tri-State Area is a very dense network of telecommunications, which developed historically in many layers. The result is what Lewis describes in his book:

To make his point, Ryan brought in oversize maps of New Jersey showing the fiber-optic networks built by telecom companies. The maps told a story: Any trading signal that originated in Lower Manhattan traveled up the West Side Highway and out the Lincoln Tunnel. Perched immediately outside the tunnel, in Weehawken, N.J., was the BATS exchange. From BATS the routes became more complicated, as they had to find their way through the clutter of the Jersey suburbs. “New Jersey is now carved up like a Thanksgiving turkey,” Ryan says.

The Flash Boys of Wall Street

One way or another, they traveled west to Secaucus, the location of the Direct Edge family of exchanges owned in part by Goldman Sachs and Citadel, and south to the Nasdaq family of exchanges in Carteret. The New York Stock Exchange, less than a mile from Katsuyama's desk, appeared to be the stock market closest to him — but Ryan's maps showed the incredible indirection of fiber-optic cable in Manhattan. "To get from Liberty Plaza to 55 Water Street, you might go through Brooklyn," he explained. "You can go 50 miles to get from Midtown to Downtown. To get from a building to a building across the street, you could travel 15 miles."

The Wolf Hunters of Wall Street [*Paywall*]

https://www.nytimes.com/2014/04/06/magazine/flash-boys-michael-lewis.html?_r=0

The Flash Boys of Wall Street

So we can see here the advantage of colocation. That is, if your company's server is located at or near the server of the Exchange you are targeting, you get a small but significant timing advantage. But there are other ways to gain a speed advantage in a complex network like this one.

We'll get back to how routing works in the Tri-State Area a bit later. Keep that map slide handy!

But you don't even need to front-run your competition if you can engage in non-transparent trading off the exchanges. This practice in arbitrage involves large blocks of securities, traded in special rooms called Dark Pools.

The Flash Boys of Wall Street

Dark Pools – Off-Exchange Trading Rooms

Dark pools are private exchanges for trading securities that are not accessible by the investing public.

Dark pools were created in order to facilitate block trading by institutional investors who did not wish to impact the markets with their large orders and obtain adverse prices for their trades.

According to the most recent SEC data, there were 64 registered Alternative Trading Systems (a.k.a.: "Dark Pools") with the SEC as of Feb. 2022.

The Flash Boys of Wall Street

Why did dark pools come into existence? Consider the options available to a large institutional investor who wanted to sell one million shares of XYZ stock before the advent of non-exchange trading. This investor could either:

Work the order through a floor trader over the course of one or two days and hope for a decent VWAP (volume-weighted average price).

Split the order up into, for example, five pieces and sell 200,000 shares per day.

Sell small amounts until a large buyer could be found who was willing to take up the full amount of the remaining shares.

The Flash Boys of Wall Street

The market impact of a sale of one million XYZ shares could still be sizable regardless of which option the investor chose since it was not possible to keep the identity or intention of the investor secret in a stock exchange transaction. With options two and three, the risk of a decline in the period while the investor was waiting to sell the remaining shares was also significant. Dark pools were one solution to these issues.

The lack of transparency actually works in the institutional investor's favor since it may result in a better-realized price than if the sale was executed on an exchange. Note that, as dark pool participants do not disclose their trading intention to the exchange before execution, there is no order book visible to the public. Trade execution details are only released to the consolidated tape after a delay.

The Flash Boys of Wall Street

While regulated dark pools are legal and regulated by the SEC, they have been subject to criticism due to their opaque nature. Because dark pool trades are not available to the public, they have sometimes been used for predatory practices by high-frequency trading firms, using tactics such as “pinging” dark pools to unearth large hidden orders and then engaging in front-running or latency arbitrage.

Dark pool operators have also been accused of misusing their dark pool data to trade against their other customers or misrepresenting the pools to their clients. According to the *The Wall Street Journal*, securities regulators have collected more than \$340 million from dark pool operators since 2011 to settle various legal allegations.

The Flash Boys of Wall Street

What Are Dark Pools in Cryptocurrency?

Cryptocurrency dark pools are similar to the pools in the stock market: they match buyers and sellers for large orders, without a public order book. However, cryptocurrency dark pools may be decentralized through the use of smart contracts. Instead of trading through a trusted intermediary, the buyers and sellers interact through a blockchain-based program that executes the trade without disclosing any confidential information about the transaction or its participants.

An Introduction to Dark Pools

<https://www.investopedia.com/articles/markets/050614/introduction-dark-pools.asp>

The Flash Boys of Wall Street

So there are both legal and highly unethical if not illegal uses for Dark Pools. In crypto blockchains and in some of the “Black Bag” derivatives markets, the problems have become only too obvious.

But in Wall Street and on the Chicago commodities exchange, there have been some reasons for using Dark Pool arbitrage, and these markets have seemed to benefit from the pricing stability these off-exchange transactions can offer. But the lack of transparency is ripe for accusations and actual occurrences of market manipulation and full-blown rigging.

So, what happened in the case of the “Flash Boys” of Wall Street? (SWITCH TO DOCUMENT – SHARE MY PICTURE)

The Flash Boys of Wall Street

According to Lewis:

The stock market really was rigged. Katsuyama often wondered how enterprising politicians and plaintiffs' lawyers and state attorneys general would respond to that realization.

Technology had collided with Wall Street in a peculiar way. It had been used to increase efficiency. But it had also been used to introduce a peculiar sort of market inefficiency. Taking advantage of loopholes in some well-meaning regulation introduced in the mid-2000s, some large amount of what Wall Street had been doing with technology was simply so someone inside the financial markets would know something that the outside world did not.

The Flash Boys of Wall Street

The same system that once gave us subprime-mortgage collateralized debt obligations no investor could possibly truly understand now gave us stock-market trades involving fractions of a penny that occurred at unsafe speeds using order types that no investor could possibly truly understand. That is why Brad Katsuyama's desire to explain things so that others would understand was so seditious. He attacked the newly automated financial system at its core, where the money was made from its incomprehensibility.

The Wolf Hunters of Wall Street

By Michael Lewis

https://www.nytimes.com/2014/04/06/magazine/flash-boys-michael-lewis.html?_r=0

[Paywall]

The Flash Boys of Wall Street

A pretty grim assessment. I do not necessarily agree with this conclusion, but it has been presented in a compelling way. In any event, the story and the book are fascinating reading. Serious questions are raised.

And it all comes down to a few milliseconds of network latency.

Videos of interviews about The Flash Boys:

<https://www.youtube.com/watch?v=7Olf2TiuAos> 1:13:18

<https://www.youtube.com/watch?v=OlobbmeO03A> 1:11:00

<https://www.pbs.org/newshour/show/flash-boys-investigates-high-frequency-traders-anticipate-wall-streets-next-move-faster> 0:08:32

The Flash Boys of Wall Street

More Videos and Interviews about The Flash Boys:

https://www.youtube.com/watch?v=x387_k963yY 0:57:43

<https://www.cbsnews.com/video/wall-street-reforms-after-rigged-trading-revealed/> 0:06:12

The Flash Boys of Wall Street

(DISCUSSION and QUESTIONS)

-- Bob Primak --

June 22, 2022

For the Lexington Computers and Technology Group (LCTG).