



The Matrix, but with money: the world of high-speed trading

Supercomputers pitted against one another in a high-stakes battle of attack and counterattack over a global network where predatory algorithms trawl the information stream, competing every millisecond to gain an informational advantage over rivals. It sounds like Hollywood fiction, but it's just an average trading day on the stock market.

By [Jon Stokes](#) | Last updated July 27, 2009 11:12 PM CT

It sounds like something out of *The Matrix*: a giant, world-spanning electronic network where high-powered machines, some of them using GPUs to gain a speed advantage, run secret, rapidly-evolving software algorithms that battle it out for profits in a high-stakes game of cat-and-mouse, attack-counterattack, that yields some \$21 billion a year for the winners and can spell ruin for the losers. Except that it's not *The Matrix*—it's the stock and commodities markets, and the fact that these markets mainly consist now of computers trading against one another has been brought closer to the public's attention by [last month's alleged theft](#) of Goldman Sachs' proprietary trading code.

The collection of computer-automated, high-speed trading technologies and techniques that are typically lumped under the heading of "high-frequency trading" (HFT) have been around for a while, but HFT has recently become heavily identified with the banking giant, Goldman Sachs, that dominates some aspects of it on the New York Stock Exchange. And as Goldman draws more media and congressional scrutiny, so will HFT. To prepare you for the high-frequency trading media onslaught, we'll take a look at HFT and at a stock market that really isn't what you thought it was.

If you look under the hood of the markets in 2009, you'll find that the trading floor has been replaced by electronic networks, the frantic, hand-signaling traders have been replaced by computer systems, and all of moves in the trader's dance—a thousand little tricks and techniques (some legal, some questionable, and some outright illegal) for taking regular advantage of speed, location, and information to generate profits—are executed hundreds of times per second, billions of times per day. And the whole enterprise is mainly powered by the same hardware from Intel, AMD, and NVIDIA, that Ars readers use for gaming.

The world that has been pulled over your eyes to blind you from the truth

Press reports of trading days that end with big gains or losses are typically accompanied by shots of a trading floor where young traders are either euphorically throwing papers into the air (up days) or staring dejectedly at a stock ticker with hand pressed to forehead, shoulders slumped in defeat (down days). These guys, you think, are "the market," and if you looked up the New York Stock Exchange (NYSE) on Investopedia you'd find nothing in the "[Stocks Basics: How Stocks Trade](#)" entry to disabuse you of this widely held notion. "The NYSE is the first type of exchange... where much of the trading is done face-to-face on a trading floor," Investopedia declares, and it goes on to provide a description of how a floor-centered, face-to-face NYSE that hasn't matched reality for about five years.



The floor of the New York Stock Exchange

Only about three percent of trading volume on the NYSE is actually carried out by means of traditional "open outcry" trading, where flesh-and-blood humans used to gather to buy and sell securities. The other 97 percent of NYSE trades are executed via electronic communication networks (ECNs), which, over the past ten years, have rapidly replaced trading floors as the main global venue for buying and selling every form of asset, derivative, and contract. So the ECNs *are* the markets in 2009, and those pit traders who pose for the cameras are mainly there for the cameras.

"Why don't you know BATS?," Bernard Donefer, a finance professor and HFT expert at CUNY's Baruch College, asked me rhetorically. "Because there's nothing to look at. It's based in Kansas; the computers are in Jersey City."

At the time that Donefer and I spoke last week, BATS was the third largest equity market in the world, behind the NYSE and NASDAQ, and it has been all-electronic since it began life in 2005. There has never been a floor that a CNBC camera crew could report from, so it's essentially invisible to the general public. The NYSE and a few other exchanges keep hang on to their trading floors "mainly for branding purposes," Donefer told me.

The ECNs offer the advantages of speed, anonymity, error minimization, and audit trails. They've also "ported" many of the problems endemic to electronic networks—security vulnerabilities, the "garbage-in, garbage out" (GIGO) problem, and the problem of technology moving too fast for lawmakers, to name just three—from the Internet to the markets. But the problems with ECNs are a topic for another day. The real issue is that when the average retail investor gets an E*Trade account and tries to play the stock market, she typically has no idea that she's going up against the market equivalent of IBM's chess grandmaster-thumping supercomputer, Deep Blue.

Not your father's stock-picking club

Experts guess that between 60 and 75 percent of the NYSE's daily trading volume is just computers trading against one another using a variety of strategies. Recent HFT investigations by Donefer, Themis Trading, and sites like Zero Hedge have brought to light a lively ecosystem of algorithms, or "algos" in the parlance, that use ECNs in different ways to make money.

What the vast majority of these algos have in common is that they are not long-term, buy-and-hold "investors" in the classic sense. Rather, they focus on executing as many trades per second as possible and on turning a small profit (often pennies or fractions of a penny) on each trade. This combination of high speed, massive volume, and razor-thin per-trade profits adds up over the course of a day, week, or year to some very large numbers.

At least two different groups, the TABB Group and FIXProtocol, estimate that high-frequency trading generated around \$20 billion in profits for the financial sector last year. Goldman Sachs accounts for some 20 percent of global high-frequency trading activity, and the bank recently had a blow-out quarter in which its HFT-heavy trading operation racked up a record number of days where profits topped \$100 million.

Goldman Sachs may be a major player in HFT, but the bank is by no means the only one. There are thousands of firms that use HFT in the market to varying degrees, with some using HFT exclusively to generate profits. There are whole funds, like those operated by Renaissance Technologies, that consist entirely of a large computer system and some PhD programmers.

These high-frequency trading platforms consist of two main components: the software algorithms, or "algos," and the underlying hardware. Let's take brief look at each part in turn.

Algorithms: like BattleBots, but with money

Different analysts divide up the world of HFT in different ways, and such division and classification efforts are very much a work in progress right now. Indeed, many would take issue with my lumping together all computer-automated trading practices under the label of HFT; most who follow the space have terms that they prefer, and, like the space itself, the language is still in flux. This being the case, I won't attempt anything like a systematic overview of HFT—instead, I'll treat the space a bit like a zoo, and I'll merely point out a few different animals.

Iceberging and predatory algos



One of the most important uses for HFT is to get the best price for very large stock orders by breaking them up into small orders of random sizes and hiding the activity from other traders, who, on sensing that a large order is in progress, might take advantage of that knowledge by making moves that would impact the stock price. Institutional traders use this practice, called "iceberging," quite frequently, and most experts agree that it contributes to the orderly functioning of the markets by reducing wide price swings.

Some categories of "predatory algos" closely monitor the markets in order to sniff out exactly these types of hidden large orders, so that the algo can trade against them. For instance, if a predatory algo detects that someone is trying to hide a large sell order for [INTC](#) by trickling it out into the market in small blocks, it might work to bid down the price of INTC just a bit so that it can pick up those blocks at a discount and then sell them for a profit when the share price floats back up to the market's earlier, nonmanipulated valuation.

Statistical arbitrageurs

Stat arb platforms don't have to be right all of the time—they only have to make money on the majority of the thousands of trades that they execute each day, in order to be very profitable..

One very popular variety of high-frequency trader is the statistical arbitrageur, or "stat arb." Stat arbs make their money by vacuuming up mountains of historical data and looking for correlations between various datapoints and asset prices. The stat arb's trading platform, which is basically a large computer system manned by programmers and financial engineers, uses those

correlations to build predictive models that take in a stream of information inputs like news reports and stock prices (Thompson Reuters sells a service that fires wire reports at very low latency to these systems), and output a rapid-fire stream of "buy" and "sell" orders for different assets.

For instance, a stat arb HFT platform might identify a direct correlation between positive news about Steve Jobs' health and increases in the price of AAPL; then the microsecond that the platform receives and processes an in-bound Reuters news packet containing a statement about Jobs' cancer-free status, it would immediately spit out a "buy" order for AAPL on the expectation that Apple stock is about to increase in price once this news becomes more widely known.

The Apple example above is a bit contrived, but it's hard to come up with a real-world example because the models and algorithms used by stat arb shops are closely guarded trade secrets. These platforms trade every asset class available on the electronic markets—from frozen pork bellies to complex structured financial products—and they constantly scour the earth looking for data that will let them spot correlations and build or refine their predictive models.

Stat arb platforms don't have to be right all of the time—they only have to make money on the majority of the thousands of trades that they execute each day in order to be very profitable. Making money on these trades requires two things: speed and good correlations. First, the stat arb's platform needs to process the information very fast and get its orders to the exchange quickly, preferably ahead of its competitors, which may be using a similar model and may get to a trade just a few microseconds later. Every fraction of a second lost is a penny or two lost to the competitor who got to the trade just before you did.

The other thing that stat arbs need is for the correlations that they've spotted to actually hold true; or, if those correlations stop working, then they need to be able to find new correlations that do work with a relatively quick turnaround. There are some periods before a major disruption when well-functioning market correlations begin to distort *en masse*; in such periods, the most recent of which was in August of 2007, the so-called "market neutral" funds that rely on these correlations to turn profits regardless of the market's overall direction can begin to tank dramatically.

Dark pools

The final animal in the HFT menagerie that I'll point out on this brief tour is the automated market maker (AMM), which is a subtype of what is often called "dark pools," or "dark liquidity." AMMs like Citadel and GETCO always stand ready to buy and sell large quantities of assets, they don't publish price quotes to other market participants via exchanges.

To find out what assets a dark pool will either sell or buy and at what price, you first have to ping it. Once you ping the pool with a request to, say, buy a specific asset, the pool will reply with the price that it's willing to sell you that asset for. You can either accept the price and complete the transaction, or turn it down and ping again later to see if the price has moved in your direction.

Dark pools, then, let traders completely sidestep normal stock and commodities exchanges in order to buy and sell assets without having to broadcast their desired price to the rest of the world. This anonymity has a number of uses, the most important of which is that it makes it easier to implement the "iceberg" strategy described above.

As with the other creatures in our HFT menagerie, the AMM makes money a little at a time, but in volume. These massive, computer-automated brokers turn a profit by making money on the bid/ask spread and by not being too exposed to price movements in any one asset class.

Dark pools have come under fire precisely because of the secrecy and anonymity that they're designed to enable. In most respects, these pools are the polar opposite of the "open outcry" markets of yesteryear, and that lack of transparency is raising hackles. There are accusations that dark pool owners abuse their privileged role as an intermediary in these opaque markets by somehow manipulating prices and/or increasing the bid/ask spread, but even those who aren't willing to go so far as to level accusations of illegal or unethical behavior see in this secrecy and lack of oversight the potential for massive abuse.

Hardware: game on

Because high-frequency trading is, as Richard Bookstaber has [recently described it](#), an "[arms race](#)" [where relative speed matters](#) much more than absolute speed, this market is one of the few left with a demand for raw performance at any cost. Indeed, my personal introduction to the world of HFT came in bits and pieces over the past few years via parts of briefings from the Intel, NVIDIA, AMD and their would-be competitors, all of whom have been aggressively pursuing this market.



Anecdotally, NVIDIA has been fairly successful in making inroads into the market with CUDA, despite any concerns about vendor lock-in. The amount of money potentially gained by a significant speed advantage is so high that programmer time is cheap by comparison, and the result is that HFT shops quickly began to experiment with CUDA when NVIDIA released the SDK.

Intel is likely to be another beneficiary in this space, since its Xeon parts have maintained a consistent top-end performance lead vs. AMD over the past few years. I say "likely to be,"

because HFT shops are very secretive about every aspect of their platforms. The only information that I have about the popularity of Xeon in HFT comes informally from Intel, so it doesn't count for much. I've heard that NVIDIA and CUDA are popular from multiple sources, though.

In all, it's ironic that the hardware that HFT platforms are using to battle it out over stocks, bonds, commodities, and other assets is essentially the same as the technology that PC gamers are using to play their own games with much lower stakes.

The other key ingredient to the success of any HFT platform is low network latency. The platforms are greatly helped in this regard by the fact that many exchanges will let HFT platforms pay to co-locate their servers with those of the exchange itself, so that the HFT platform can get its order in ahead of the competition. Critics contend that such co-location deals provide avenues for potential front-running of orders, in which an HFT platform gets an advance peek at an incoming order that will move a stock's price in a specific direction, and then uses that knowledge to make a quick bet on impending the price move.

Too much, too fast?

Apart from the issues of transparency and oversight raised by the HFT approaches described above, there's also the possibility that HFT, with all of its enormous speed and complete automation, poses a larger systemic risk to our markets

I mentioned earlier in this article that high-frequency trading was "estimated" to account for between 60 and 75 percent of all available market volume. This number, which you might think would be important to know, is only one of a number of survey-based, ballpark estimates; the real numbers aren't knowable because algo trades aren't marked as such. In other words, we have no way to tell how much of the current stock market activity—both prices and volume—is the result of computers trading against each other in the manner described above.

It's also not clear whether all of this computerized buying and selling is actually good for the markets and for society as a whole. Couldn't we as a society better spend all of this money, computer power, and PhD brainpower on, say, coming up with a fossil fuel replacement? Supporters of HFT respond that their platforms provide much-needed market liquidity. They argue that, without HFT, there may not be enough buyers or sellers for a particular asset, so the market in that asset just stops functioning smoothly.

Not everyone is convinced that liquidity is worth the attendant risks of HFT, which are very difficult to quantify when you're looking at HFT's potential impact on the market as a whole.

Apart from the issues of transparency and oversight raised by the HFT approaches described above, there's also the possibility that HFT, with all of its enormous speed and complete automation, poses a larger systemic risk to our markets.

At the back of everyone's mind is the 1987 program trading crash, described by Richard Bookstaber in [*A Demon of our Own Design*](#). In the run-up to October of 1987, all of the major

market participants had been using essentially the same computer-automated algorithm to hedge their portfolio risk. On Black Monday (10/19/1987), all of the portfolio insurance programs started dumping assets in lock-step, in response to a particular set of inputs. This synchronized selling begat more synchronized selling, and by the time this giant, market-sized feedback loop was shut down by the closing bell, the Dow had lost almost 23 percent of its value in a single day.

Most of the debate around HFT is between those who think that a similar crash could not only happen again, but could be many times worse because the aforementioned increases in speed and trading volume, and those who insist that we don't yet know enough to make that call. It could be that this fast-moving system as a whole could quickly and dramatically fail in some unforeseen way, due to a combination of an external shock and unseen internal fragility; or, it could be redundant and robust enough to keep humming along in the face of anything we (or Mother Nature) throw at it.

Either way, though, HFT's combination of speed, volume, secrecy, and lack of human oversight and intervention worries even those who trust the human players not use their machines to cheat at the game.

Further reading

- [Goldman's secret sauce could be loose online; markets beware](#)
- [For Whom the Bell Tolls: The Demise of Exchange Trading Floors and the Growth of ECNs](#)
- [FINalternatives Survey: High-Frequency Trading Has a Bright Future](#)
- [Risk Management and Electronic Trading](#)
- [Stat Arbitrageurs: Merchants of Volatility](#)
- [The Real Story of Trading Software Espionage](#)
- [Toxic Equity Trading Order Flow On Wall Street](#) [PDF]