

“Clearly one of the most innovative and transformational CIOs ever”

☰ Menu

Wall Street’s Quest To Process Data At The Speed Of Light

👤 Dr. Steve 📁 Innovation 🕒 21 Apr 2007 18 Feb 2018 ☰ 11 Minutes

Financial firms use physical proximity to overcome the technical barriers of data latency

By Richard Martin, *Information Week*

21 April 2007


URL: <http://www.informationweek.com/showArticle.jhtml;jsessionid=WKLLSLTGL0NVQQSNDLRSKH0CJUNN2JVN?articleID=199200297&queryText=rubinow>
(<http://www.informationweek.com/showArticle.jhtml;jsessionid=WKLLSLTGL0NVQQSNDLRSKH0CJUNN2JVN?articleID=199200297&queryText=rubinow>)

When the New York Stock Exchange celebrated its Euronext NV merger at the closing bell on April 4, the trading floor erupted in boos and jeers—and the ringing of cowbells.

The cowbells came from NYSE officials hoping to drown out the catcalls of traders, who see the merger as yet another threat to their livelihoods. With the spread of computerized trading—electronic trading now makes up 60% to 70% of the daily volume on the NYSE and algorithmic trading close to half of that—the manic floor traders look to be headed the way of the exchange’s in-house barber (long gone). In the last year, the number of traders on the NYSE has dropped by a quarter.

Executing complex strategies based on arcane mathematical formulas, algorithmic trading systems generate thousands of buy and sell orders every second, many of which are canceled and overridden by subsequent orders, sometimes only a few seconds apart. The goal of these computer (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=computer&x=&y=>) traders is to profit from minute, fleeting price anomalies and to mask their intentions via “time-slicing,” or carving huge orders into smaller batches so as not to move the market.

A 1-millisecond advantage in trading applications can be worth \$100 million a year to a major brokerage firm, by one estimate. The fastest systems, running from traders’ desks to exchange data centers, can execute (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=execute&x=&y=>) transactions in a few milliseconds—so fast, in fact, that the physical distance between two computers processing (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=processing&x=&y=>) a transaction can slow down how fast it happens. This problem is called data latency—delays measured in split seconds. To overcome it, many high-frequency algorithmic traders are moving their systems as close to the Wall Street exchanges as possible.

 Trying to stay ahead of electronic upstarts, the NYSE's Rubinow plots an all-automated future

Twiddling the knobs on these fantastically complicated, blindingly fast systems are human traders known as quant jocks (for quantitative trading tactics) and stat arbs (for statistical arbitrageurs). They work for major brokerages such as Credit Suisse and Merrill Lynch, or for little-known, publicity-shy hedge funds such as SAC Capital and Renaissance Technologies (both of which declined to comment for this story). They're competing to shave fractions of seconds off transaction times in order to pinpoint share prices.

Wall Street's quest for speed is not only putting floor traders out of work but also opening up space for new alternative exchanges and electronic communications (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=communications&x=&y=>) networks that compete with the established stock markets. Electronic trading has reduced overall volatility in the equities markets, because volatility is a product of herd buying or selling, and electronic trading—responding instantaneously to tiny price fluctuations—tends to smooth out such mass behavior. And it has provided established exchanges with new revenue opportunities, such as co-location services for companies that wish to place their servers in direct physical proximity to the exchanges' systems. Electronic trading also has created opportunities for a new class of vendors—execution services firms and systems integrators promising the fastest possible transaction times.

At its most abstract level, the data-latency race represents the spear point of the global movement to eradicate barriers—geographic, technical, psychological—to fair and transparent markets. “Any fair market is going to select the best price from the buyer or seller who gets their order in there first,” says Alistair Brown, founder of Lime Brokerage, one of the new-school broker-dealers, which uses customized Linux (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=Linux&x=&y=>) servers to trade some 200 million shares a day. “At that point, speed definitely becomes an issue. If everyone has access to the same information, when the market moves, you want to be first. The people who are too slow are going to get left behind.”

That reality extends beyond Wall Street. While no other industry depends so much on split-second response times and high throughput between many different agents, the need for speed is growing in other sectors. Rich-media companies (those involved in video (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=video&x=&y=>), postproduction, digital animation, broadcasting, and Web 2.0), oil and gas producers, big retail chains, research institutions—they're all finding that rapid access to data is increasingly a competitive differentiator.

“A lot of different industries are trying to relieve input/output bottlenecks even if they're not operating at the sub-100-millisecond level like financial traders,” says Josh Goldstein, VP of product marketing at DataDirect Networks, a provider of high-end storage (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=storage&x=&y=>) systems, including a highly optimized algorithm (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=algorithm&x=&y=>) called DirectRAID for storing and instantly retrieving large files. “They need faster times to gain insight on their data, and they can see how that's driving their competitive advantage.”

The high-speed data transfer segment of the storage industry is growing 20% to 25% a year, Goldstein says, while DDN's own business is growing even faster. In that sense, Wall Street's latency-reduction efforts are helping drive the development of network and storage technologies—tiered storage that deposits different types of data according to how often it's accessed, advanced interconnect technologies

like InfiniBand, and high-speed, front-end controllers that provide data access to multiple computers—that will spread to other industries, for everything from 3-D seismic exploration information to archived video.

VALUE IN MILLISECONDS

On the New Jersey side of the Lincoln Tunnel, in an anonymous three-story building, is one of the financial world's most important data centers. Pushing the door bell at the unmarked main entrance won't get you inside—it's merely a facade. The real entrance is harder to find.

The servers for five electronic exchanges are located in this data center, along with computers belonging to dozens of trading firms. Run by [hosting](http://www.techweb.com/encyclopedia/defineterm.jhtml?term=hosting&x=&y=) (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=hosting&x=&y=>) company Savvis, the Weehawken facility is home to some of the most advanced trading technology anywhere. Much of Savvis' growth can be traced to the spread of what's known as direct market access. In the past, traders used consolidated feeds—market data updates such as those provided by Reuters and Thomson. Distributing those feeds, however, could take up to 500 milliseconds, far too long for today's automated trading.

"Now you're seeing a lot of the market data providers and vendors who have direct exchange-feed connectivity," says Varghese Thomas, Savvis' VP of financial markets. Savvis provides connectivity from the exchange directly to the [client](http://www.techweb.com/encyclopedia/defineterm.jhtml?term=client&x=&y=) (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=client&x=&y=>) without having to go through a consolidated system.

The exchanges themselves also are profiting from the demand for [server](http://www.techweb.com/encyclopedia/defineterm.jhtml?term=server&x=&y=) (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=server&x=&y=>) space in physical proximity to the markets. Even on the fastest networks, it takes 7 milliseconds for data to travel between the New York markets and Chicago-based servers and 35 milliseconds between the West and East coasts. Many broker-dealers and execution-services firms are paying premiums to place their servers inside the data centers of Nasdaq and the NYSE.

About 100 firms now co-locate their servers with Nasdaq's, says Brian Hyndman, Nasdaq's senior VP of transaction services, at a going rate of about \$3,500 per rack per month. Nasdaq has seen 25% annual increases in co-location the last two years.

Physical co-location eliminates the unavoidable time lags inherent in even the fastest wide area networks. Servers in shared data centers typically are connected via [Gigabit](http://www.techweb.com/encyclopedia/defineterm.jhtml?term=Gigabit&x=&y=) (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=Gigabit&x=&y=>) Ethernet, with the ultrahigh-speed switching fabric called InfiniBand increasingly used for the same purpose, says Yaron Haviv, CTO at Voltaire, a supplier of systems that can achieve latencies of less than a microsecond, or 1 millionth of a second.

Later this year, Nasdaq will shut down its [data center](http://www.techweb.com/encyclopedia/defineterm.jhtml?term=data%20center&x=&y=) (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=data%20center&x=&y=>) in Trumbull, Conn., and move all operations to one opened last year in New Jersey, with a [backup](http://www.techweb.com/encyclopedia/defineterm.jhtml?term=backup&x=&y=) (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=backup&x=&y=>) in the mid-Atlantic region, Hyndman says. (Trading firms and exchanges are reluctant to disclose the exact locations of their data centers.)

The NYSE will begin reducing its 10 data centers, including those associated with Euronext, to two in the next couple of years, says CTO Steve Rubinow. Co-location, Rubinow says, not only guarantees fast transactions, but also predictable ones. "If you've got some trades going through at 10 milliseconds and some at 1 millisecond, that's a problem," he says. "Our customers don't like variance."

One of the biggest co-location customers is Credit Suisse, which handles about 10% of all U.S. equity trades daily and which helped pioneer black-box trading systems with exotic algorithms that go by monikers like Sniper, Guerrilla, and Inline. Credit Suisse maintains Sun and Egenera blade (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=blade&x=&y=>) servers, some running Linux and some Windows, in all the major U.S. markets, says Guy Cirillo, manager of global sales channels for Credit Suisse's Advanced Execution Services unit, which serves major hedge funds and other buy-side clients. The AES (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=AES&x=&y=>) trading engine in Credit Suisse's Manhattan headquarters is replicated in London, Hong Kong, and Tokyo.

Guaranteed transaction times for AES clients—from the time the order is received on the Credit Suisse system until it gets an acknowledgment from the exchange, electronic communications network, or "crossing network"—has dropped from 15 milliseconds to 8 in the last year, Cirillo says. Total execution time also includes any delays within the exchange or "liquidity point" itself, a latency (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=latency&x=&y=>), variable over which Credit Suisse has no control.

"That response time is something the ECNs and the exchanges compete on as well," Cirillo says. "Their latency, their turnaround time, and their infrastructure (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=infrastructure&x=&y=>) are all part of the electronic game."

Founded in 1971 as the first all-electronic equities market, Nasdaq has made a series of acquisitions of upstart ECNs, buying the Brut and INET (<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=INET&x=&y=>) electronic trading systems in the last few years. Nasdaq is now shifting its legacy trading platform—called Supermontage, based on Hewlett-Packard, Stratus, and Dell servers—over to INET's hyper-rapid system. The NYSE, meanwhile, purchased Chicago-based Archipelago in April 2005.

Hyndman, who oversaw the acquisition of Brut and INET, says they've allowed Nasdaq to offer the fastest possible transaction times. In migrating to INET, "we've reduced latency from 10 milliseconds down to 1," he says, referring to the time it takes once an order is placed on Nasdaq's system to acknowledge it electronically. "It gives customers the speed they're looking for, with the tremendous throughput, reliability, and stability they're looking for in a matching engine."

Also driving Wall Street's electronic revolution has been government regulation, specifically the Regulation National Market System, or Reg NMS. Intended to modernize and standardize equity markets, Reg NMS has several provisions, the most significant and controversial of which is the trade-through rule, which essentially states that all customer orders must be routed to the exchange with the best price at the moment the order is entered. One consequence of that mandate, which took effect for exchanges in early March and will be applied to broker-dealers later this year, is that holdout manual exchanges, such as the NYSE and regional exchanges in Philadelphia, Boston, and elsewhere, must finally go electronic. In late February, the NYSE finished rolling out its Hybrid Market system, which integrates aspects of the auction market with automated trading.

Traders are being forced to rely on automated systems more and more, according to a report from Wall Street consulting firm the Tabb Group titled "Trading At Light Speed." And they're using automation "not only to filter the data for them, but increasingly to process that data, interpret it, and drive automated trading decisions," according to the report.

Many observers, especially high-frequency traders and new-fashioned broker-dealers such as Lime, see even hybrid systems such as the NYSE's as a stopgap. Soon, they argue, all markets will be completely electronic, and transaction times will be the ultimate arbiter of who gets the best trade at the best price.

"Computers are just so much faster and more efficient that there's no point in doing things the old-fashioned way," says Mark Akass, CTO for BT's global financial services unit. "Computers make both the decision process and the execution so much faster that eventually everything will be done electronically."

That doesn't mean that they're infallible. The sequence of events that followed the sharp decline in the Chinese stock market in late February was compounded by a computing error that made it appear that the Dow Jones industrial average had dropped 178 points in a minute. Automated trading systems responded, and the Dow plunged 416 points, or 3.3%, on the day. The NYSE was forced to suspend electronic trading that afternoon.

OUT WITH THE OLD

Regardless, high-speed, automated trading over electronic networks eventually will make the traditional exchanges obsolete, or nearly so, predicts David Cummings, CEO and founder of BATS Trading. Founded in June 2005, BATS (for Better Automated Trading System) was built from the ground up to take advantage of the new black-box trading schemes and, says Cummings, to compete with Nasdaq, which he regards the way a fighter jet pilot might look at a biplane.

Tough-talking Cummings is perhaps the leading proponent of the theory that the push to reduce data latency represents not just an imperative for Wall Street traders but a new era for Wall Street itself. "Nasdaq spent \$1.6 billion to buy out INET and Brut," pronounces Cummings. "I told my investors, 'For \$30 million or \$50 million we can rebuild it.' And we are."

BATS's trading system, based on multicore HP servers running proprietary software, initially cost all of \$2 million, and it averaged response times of less than a millisecond, Cummings says. That kind of performance earned BATS the top spot from research firm Celent in its latest industry-benchmark rankings of executions speeds, above Nasdaq's trading engine. And, Cummings says, it will get BATS to 1 billion shares traded per day by the end of this year, from just over 300 million in February.

BATS isn't the only trading network to spring up in response to the new techniques and technologies. Direct Edge, CBSX (an offshoot of the Chicago Board Options Exchange), and Liquidnet (which last month acquired Miletus Trading, a broker-dealer for algorithmic trading) all have appeared in the last few years.

Meanwhile, established providers of market data have introduced products to serve high-frequency black-box traders. Responding to the direct-feed explosion, Reuters last year rolled out Reuters Data Feed Direct, a managed service for high-end traders that offers latency of less than 1 millisecond and a "full liquidity view of the market," says Kirsti Suutari, global business manager for algorithmic trading in Reuters' enterprise trading division.

“The use of machines begets the use of machines,” notes Suutari. “As the market becomes more electronic and more appropriate for machine-based trading, we’ve seen this huge increase in the market data update rate. We needed to create a product more designed for that kind of volume of data flow.”

The stock market’s update rate—the real-time data flow on prices and transactions—has now reached something like 83,000 messages per second, Suutari says.

Also offering a new avenue for high-frequency trading is BT Radianz, which resulted from BT’s 2005 acquisition of trading-platform provider Radianz. Launched in March, its Ultra Access service provides sub-1 millisecond order-routing services between traders and exchanges in the New York area. That’s as fast as you can get, says BT’s CTO Akass. “After you get below 1 millisecond, you get to the speed-of-light limitation,” he says. “There’s not a lot you can do in terms of getting faster.”

BEYOND WARP SPEED

Therein lies the rub for ultrafast trading: Once you hit physical limits to data-transmission speeds, where do you go from there? “If anybody knows how to get a signal transmitted faster than the speed of light, I’d like talk with them,” says Cummings.

There are two schools of thought on this issue. One is that traders, exchanges, and brokers must shave latency from other parts of the system—in the applications they use, for instance—and that the race will continue.

The other is that latency will cease to be an issue once everyone has access to the same trading infrastructure and that other, older-school elements of the business such as customer service and market savvy will, once again, become the differentiators. “Shortly, we’ll be talking micro- versus milliseconds, and at that point speed will probably have less and less relevance,” says Brown of Lime Brokerage. “Once you’ve got half a dozen systems that can all handle that kind of throughput, then you have to distinguish yourself somewhere else.”

For now, though, the quest for speed continues. And Wall Street’s transformation goes on, at faster and faster rates. “Five years ago we were talking seconds, now we’re into the milliseconds,” says BATS CEO Cummings. “Five years from now we’ll probably be measuring latency in microseconds.”

Millionths of a second? If there’s a will—and a multimillion-dollar stock trade at stake—there’s bound to be a way.

—With Elena Malykhina

Published by Dr. Steve

[View all posts by Dr. Steve](#)

