Market Structure Insights

Andrew Upward
Head of Market Structure
aupward@weedenco.com
203.861.7618

IEX: An Intriguing Wrinkle in the National Market System

EXECUTIVE SUMMARY

IEX's speed bump is a clever innovation that protects hidden liquidity providers from trading at stale prices. The benefits to the liquidity provider spill over to the liquidity-taking side of the trade, which is where we think institutional investors will find the most value. We applaud this effort to enhance market structure for the benefit of institutional investors. Displayed liquidity on IEX is a different story. It's tougher to figure out what types of traders will make use of it, and in what scenarios. Granting trade-through protection to these orders presents costs that must be weighed against the benefits of allowing IEX to compete as a full-fledged exchange. Valid counter-arguments notwithstanding, we feel that the costs outweigh the benefits, and that the rules suggesting that IEX's quotes should be denied trade-through protection should be upheld.

INTRODUCTION

No topic stirs up as much passion and curiosity among our clients as IEX, the Alternative Trading System applying to become the 13th US stock exchange. Amid all the talk about the need for a holistic review of the market's rules and structure, the folks at IEX went out and built a business that strikes at the heart of many of the practices that frustrate large swaths of the buy- and sell-sides of the industry. In just two and a half years, thanks to the good will they've generated among this buy- and sell-side cohort – and thanks to the publicity boost from Michael Lewis' best-selling book *Flash Boys* – it's grown to become the second-largest ATS in the country.

Now that IEX is attempting to bring its model into the exchange realm, some of Wall Street's biggest players have strenuously objected in a series of public comment letters. They argue that as an exchange, IEX would violate some of the rules and conventions that underpin the National Market System – the web of exchanges, ATSs and broker-dealers that together facilitate all trading of US listed securities. IEX vehemently disagrees, and adds that if it does violate any rules, they're rules that other exchanges violate too.

In this report, we first consider how the IEX model would work if the Securities and Exchange Commission approves its exchange application "as is," making allowance for the impact of its recently amended routing scheme. We then discuss the rules and conventions that IEX would violate as a full-fledged exchange, and explore an idea for a compromise that we think has real merit: approving IEX's application but denying its quotes trade-through protection. In the appendix at the end of the report, we illustrate the IEX basics in a series of diagrams: what is the "speed bump," how does it work, and how does it add value to IEX's hidden peg orders and router.

PART I: THE EXCHANGE

In this section, we speculate about the types of liquidity that IEX's three main components – the dark pool, the displayed pool and the router – are likely to attract or repel when it becomes an exchange. We assume throughout the section that the SEC is going to approve the venue's exchange application in its current form. IEX is proposing to operate similarly to how it operates today as an ATS, so we don't believe there will be any major surprises.

The router will look different, however. On February 29, IEX announced that it would be making an adjustment to its order-handling procedures for routable orders. This adjustment, which we explain below and diagram in the appendix, will likely reduce the flow of big marketable orders into IEX. And that may have knock-on effects that constrain the venue's potential market share.

Still, the pool of hidden peg liquidity has intrinsic value and should continue to attract a healthy balance of liquidity providers and takers. The pool of displayed liquidity is a wild card. Our best guess is that it will garner more interest from brokers than it does today, but that the HFT community will be slow to embrace it.

THE EXISTING ROUTING SCHEME MAKES IEX BIGGER

Any router is mechanically capable of performing the tasks that IEX's router performs today. What sets IEX's router apart is its ability to operate in the microseconds following an execution on IEX without the chaos – cancels, modifications, trades, re-pricings, etc. – that often accompanies an execution. We illustrate this chaos in Figures 9, 10 and 11 in the appendix, and in Figures 12, 13 and 14 we show how the speed bump keeps the chaos at bay by creating a temporal buffer between the IEX execution and subsequent executions at away markets.

No other exchange router can operate with such a buffer – not in relation to executions taking place on IEX certainly, and not in relation to executions on its own book either. No other ATS-operated router that we know of operates this way. Broker-dealer-operated routers that don't work in tandem with an ATS *can* schedule individual routes to arrive at away markets at around the same time, but they can't offer such routes on the heels of a potential fill in an ATS.

This means that IEX alone offers what is essentially a package deal for aggressive liquidity consumers: a free option to consume liquidity on the IEX book before going out to undisturbed away markets with any unexecuted portion of the order. To the extent that this unique package succeeds in attracting aggressive order flow, the IEX liquidity pool gets a valuable look at what are effectively "parent orders" – orders that have not yet been broken up into smaller "child orders" for routing to away markets.

For an agency broker looking for block-sized fills, providing liquidity in a pool known to attract bigger, parent-sized marketable orders is a no-brainer. For an HFT market maker, intercepting a big order before it gets broken up and sprayed to away markets could be more attractive than being on the receiving end of the spray at those away markets – especially if the intercepting is done with a speed bump-protected hidden order (more on this below). So not only does the presence of bigger marketable orders logically imply bigger fills and thus a bigger market share for IEX, the knowledge that those bigger

orders are coming through incentivizes providers to post in bigger size. As the saying goes, liquidity begets liquidity.

THE NEW ROUTING SCHEME WILL LIMIT MARKET SHARE

As rival exchanges and some trading firms have pointed out, the unique value that the speed bump adds to IEX's routing capabilities would seem to qualify as "unfair discrimination between customers, issuers, brokers, or dealers." Specifically, IEX would seem to be giving its own router, which will be registered as a broker-dealer once IEX becomes an exchange, an unfair advantage over other broker-dealers. The language cited above is taken from Section 6(b)(5) of the Securities Exchange Act of 1934, and any exchange whose rules are found to permit such discrimination "shall not be registered as a national securities exchange."

We'll never know whether the SEC would have interpreted IEX's routing advantage as unfair discrimination, because IEX recently announced that it is changing how it handles routable orders. From what we can glean from publicly available documentation, once a routable order has traversed the speed bump, it will be sent directly to the router and not to the matching engine, which is where it is sent today. From there, the router will "look" at the market data it is receiving from all thirteen exchanges and decide on a schedule for the child orders it will send to IEX and away markets. The market data it receives from IEX itself is the top-of-book feed, TOPS, and it is delayed by 350 µs by a speed bump that sits between the router and the matching engine. That speed bump also delays by 350 µs any orders sent to the IEX matching engine by the IEX router. In Figures 15-18, we illustrate how the new system architecture will work in the same "urgent liquidity demand" scenario used in Figures 9-14.

As IEX states in its most-recent comment letter, the new configuration *will* protect the sender of the routable order from quote fading, just as the current configuration does. And the router itself will be just as effective as it is today. The value that the router adds to the IEX liquidity pool, however, will diminish. Orders resting on the IEX book will pair off against a child order, not the bigger parent order. And the child order will only be as big as IEX's total displayed interest, plus any shares that IEX tacks on in search of hidden liquidity when there aren't enough displayed shares across all markets to satisfy the full size of the marketable order.

We can only guess to what extent the new configuration will limit IEX's market share. Stats from IEX's website show that for every 100 shares IEX matches on its own book, it routes about 30 residual shares to away markets. These numbers don't tell us how much of the matched volume came from routable orders, unfortunately. It could be that senders of routable orders are getting nothing done on IEX before their orders are routed out, and the matched volume is all from non-routable orders. Or it could be that the senders of routable orders are getting massive amounts done on IEX, and in fact all of its matched volume is from these orders.

Obviously, the truth lies somewhere between these two extremes. We also need to bear in mind that some brokers will continue to use the router for its ability to schedule child orders for maximum displayed liquidity capture. Our gut tells us that the market share limitation will be significant, but not severe. It certainly won't kill the venue.

HIDDEN PEGS SEGMENT THE MARKET

IEX "the exchange" is planning to operate its dark pool in much the same way that IEX "the ATS" operates it today: as a pool of hidden peg orders that it re-prices using the other exchanges' proprietary data feeds. As we illustrate in Figures 6-8, once a hidden peg order traverses the speed bump and posts to the book, it can't be accessed if an NBBO change has already happened or is imminent. This feature is friendly to the liquidity provider, who is always going to be buying at a stable bid price or selling at a stable offer price. Conversely, it discriminates against the liquidity taker who wants to trade but cannot do so – at least, not in those fleeting moments just before or after a quote change.

This is segmentation. It starts with the buy-side buying into the IEX value proposition and ethos and requesting that the sell-side send orders IEX's way. That, along with the attractiveness of the router, primes the pump with a steady stream of marketable and midpoint-pegged order flow. Those flows encourage brokers to try to provide liquidity passively and to also treat IEX as a midpoint crossing engine. But HFT market makers are also part of this story. They're drawn to any venue that offers a chance to interact with marketable flow early in its lifespan, and the speed bump sweetens the deal by ensuring that they only trade when there's a good chance of making money.

Furthermore, users of IEX's discretionary peg order – a hidden order that pegs to the near side of the NBBO but has discretion to execute at prices at or between the near side and the midpoint – will be restricted to executing only at the near side of the NBBO when IEX determines that the market is at risk of moving against the user. This feature, which IEX calls Crumbling Quote protection, empowers traders to provide liquidity at more aggressive prices by protecting them from trading at those prices when doing so is likely, in IEX's view, to produce a losing trade from a short-term P&L perspective.

In addition, the fact that the discretionary peg order posts to the near side with discretion to execute up to the midpoint – as opposed to simply posting at the midpoint – means that the user receives the price improvement when the incoming order is priced to cross the spread. These two features – Crumbling Quote protection and discretion to the midpoint – make the discretionary peg order type very attractive to liquidity providers.

While we do believe that agency brokers and their buy-side clients derive value from the pool on the provide side, they must share that value with a hungry HFT market-making community. The advantage that the speed bump confers upon the hidden peg order is decidedly short-term: it protects against the trade looking bad hundreds of microseconds from now. The advantage it confers is also predicated on the trader being indifferent to whether or not she trades: there's no guarantee that, having avoided a trade at a "bad" or "stale" price, she'll be able to get the trade off at the new, better price. Institutional investors' holding periods are measured in weeks, months or years, and they can't always be indifferent to whether or not they trade. So even though the hidden peg order has been heavily marketed to institutions, and is useful to them, it is arguably better suited to HFT market makers. As such, we suspect that those market makers will come to dominate the provide side – if they don't already.

We'd be remiss, meanwhile, if we didn't highlight the advantages that accrue to the liquidity taker. We stated earlier that the speed bump discriminates against the liquidity taker, and that's true – but only for a trader that attempts to take liquidity just before or after the NBBO moves in their favor. Traders seeking liquidity at any other time will not only have an opportunity to cross against natural midpoint

and passive liquidity providers, they'll also find HFT market makers emboldened by the speed bump's protection to provide more and/or better-priced liquidity than they might otherwise feel comfortable providing on a non-speed bump market. As most retail brokers and wholesalers can attest, this is the beauty of segmentation: when it's done well, value accrues to both sides of the trade.

We're confident that the value the IEX dark pool is bringing to the marketplace through segmentation will persist after it becomes an exchange. That said, we think institutional investors and their agency brokers won't always be able to derive value from it as passive, non-block liquidity providers, and that the most common interaction will be an agency broker taking liquidity from an HFT market maker.

We wonder too if the buy-side will continue to use the venue as a block crossing engine, and if so, for how long. The block-crossing space is competitive, and there's nothing inherent in IEX's model that we can see that gives it a concrete advantage over other pools. That will be especially true once IEX changes its order-handling procedures for routable orders, and the flow of bigger marketable orders that currently pass through the book on their way to the router is cut off. Of course, block-crossing is a game of perception: as long as traders perceive it to be a venue for blocks, they'll keep sending their block orders there and block trades will materialize.
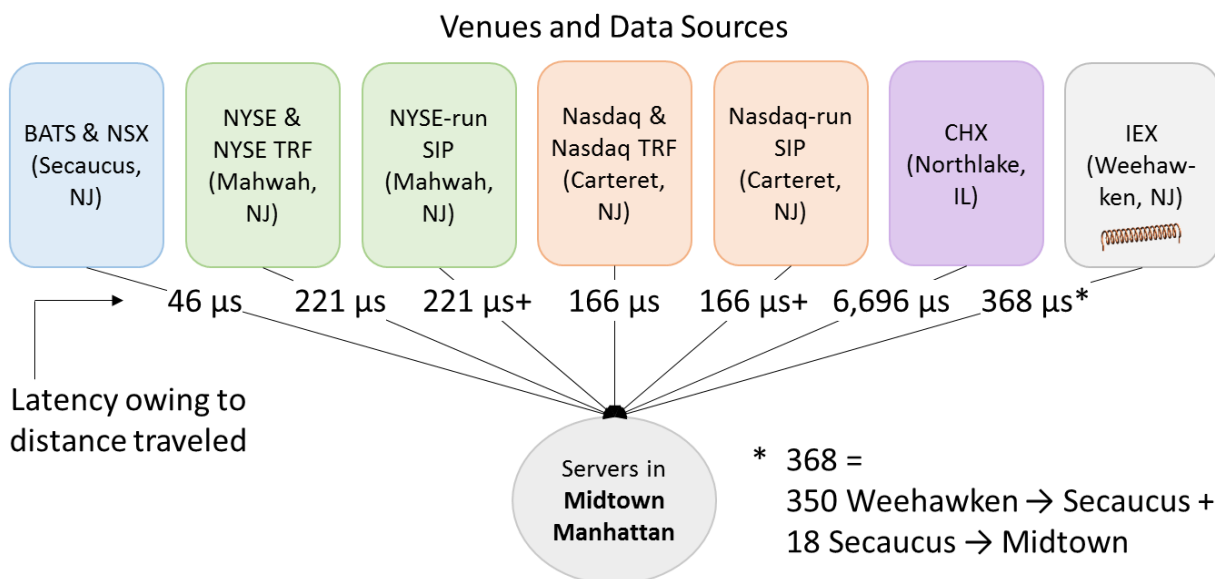
THE DISPLAYED POOL: NOT HFT-FRIENDLY

Like a hidden peg order, a displayed limit order must traverse the speed bump on its way into the IEX matching engine. Unlike a hidden peg order, once it arrives and is posted to the book, any subsequent modification or cancelation of that order must be originated by the trader and must also traverse the speed bump. In other words, IEX doesn't re-price the order on the trader's behalf based on its low-latency view of away-market quotes, as it does for its hidden peg orders.[1]

The result will likely be a displayed-liquidity pool that looks and feels much like other displayed pools to agency brokers, but that looks and feels agonizingly slow to HFT firms (in this section, we use "agency brokers" as shorthand for firms that don't co-locate their servers in multiple data centers, and "HFT firms" as shorthand for firms that do).

First, consider the agency broker's experience. If its servers were in Times Square in Midtown Manhattan, its latency experience would look something like this:

---

[1] Note that IEX *does* re-price displayed orders that would lock or cross an away market upon entry, and subsequently re-prices those orders so that they remain one tick less than the NBO (for buys) or one tick more than the NBB (for sells). But this form of re-pricing is fundamentally different than IEX's re-pricing of hidden peg orders. Also note that IEX offers non-pegged hidden limit orders that can only be modified or canceled by the trader, and not by the exchange – just like IEX's displayed orders.

Venues and Data Sources

| BATS & NSX (Secaucus, NJ) | NYSE & NYSE TRF (Mahwah, NJ) | NYSE-run SIP (Mahwah, NJ) | Nasdaq & Nasdaq TRF (Carteret, NJ) | Nasdaq-run SIP (Carteret, NJ) | CHX (Northlake, IL) | IEX (Weehawken, NJ) |

46 µs   221 µs   221 µs+   166 µs   166 µs+   6,696 µs   368 µs*

Latency owing to distance traveled

Servers in **Midtown Manhattan**

* 368 = 350 Weehawken → Secaucus + 18 Secaucus → Midtown

To arrive at these distance-driven latencies, we simply converted the "as the crow flies" distance between each pair of cities into time at the rate implied by IEX's speed bump (i.e. 350 microseconds per 38 miles). Depending on what kind of technology a firm is using, actual latencies could be higher or lower than the ones we've come up with here.

As the diagram shows, the 368 µs our agency broker has to wait before its order arrives at the IEX matching engine, and the 368 µs it has to wait before receiving market data from IEX, is not too dissimilar from the latencies to and from the other big exchanges. If our broker is relying solely on feeds from the Securities Information Processors ("SIPs"), it would receive IEX quote and trade updates around the same time that it would receive updates from Bats via the SIPs (this isn't shown in the diagram). All in all, speed bump notwithstanding, the experience of sending and modifying and canceling displayed orders on IEX will feel very similar to the broker's experience at other exchanges.

Now let's consider an HFT firm's perspective. If the firm is co-located in all three major data centers – Secaucus, Mahwah and Carteret – it will have near-zero latency to all the major exchanges. There may be scenarios or strategies for which consolidating data from the three locations is necessary, in which case the Mahwah-to-Carteret latency (313 µs using our speed bump-based conversion factor) would serve as the firm's minimum latency. Put differently, from the perspective of the firm's Mahwah servers, NYSE data will be real-time but Nasdaq and Bats data will always be slightly stale. From the perspective of its Secaucus servers, Bats data will be real-time but NYSE and Nasdaq data will always be stale, and from the perspective of its Carteret servers, Nasdaq data will be real-time but NYSE and Bats data will be stale. And there's no way for the firm to have the best of all worlds, no way to patch together and "see" all three feeds in real-time at the *same* time.

Regardless of whether the firm's strategy relies on consolidated data, the latency from IEX's speed bump could be problematic. Let's say that the HFT firm is offering stock at the same price on IEX, Nasdaq and Bats, and its offer on Nasdaq gets lifted. Its next move is to try to raise the price of its offers on IEX and Bats to avoid unwanted exposure. It routes a cancel/replace message to Bats in Secaucus from its

servers in Carteret, and within about 150 μs the message has arrived and Bats has re-priced the order. At the same time, it routes a cancel/replace message to IEX's point-of-presence in Secaucus, and within about 150 μs it enters IEX's system architecture and begins to traverse the speed bump en route to the matching engine.

Any order message that arrives at the front door of IEX's point-of-presence in Secaucus after the HFT firm's cancel/replace message has gone through it would trail the HFT firm's message as both race around and around the coiled fiber that is the speed bump. So as long as the HFT firm gets its message to the front door before any marketable buy orders, the HFT firm will successfully raise the price of its offer before it can be traded against. In this way, the HFT firm's displayed-order experience on IEX would be similar to its displayed-order experience on every other exchange.

The only problem for the HFT firm is that for a period of time that could be as long as 850 μs in our example – the 150 μs it took to get the cancel/replace message from Carteret to the IEX "front door" in Secaucus, plus the 350 μs it would take for the cancel/replace message to travel to the IEX matching engine, plus the 350 μs it would take for IEX to send an acknowledgement back – it doesn't know if it has succeeded or failed in getting its offer out of the way. That 850 μs compares unfavorably to the 150 μs it took to find out the status of its offer on Bats.

Any time an HFT firm has a limit order in the marketplace and doesn't know its status, that firm is uncertain of its exposure and thus cannot take action to reduce it. Granted, we're talking about extremely small windows of uncertainty here – hundreds of microseconds – and of course brokers that don't co-locate their servers live with these windows of uncertainty today on every order. Indeed, even the fastest, most-sophisticated HFT firms that are co-located in every major data center currently endure some uncertainty when they need to wait for information from one venue before acting on another. And that's not going to change – unless the fragmented market model itself changes.

These caveats aside, we suspect that many HFT firms won't want to tolerate the added uncertainty that comes from not knowing the status of their IEX displayed limit orders at crucial moments. We also note that IEX's proposed rule book seems to rule out "Day Intermarket Sweep Orders" and makes no mention of a "post-only" modifier for displayed limit orders. These are powerful order types that all of the incumbent exchanges offer, and that HFT firms use to exert tighter control over their orders. The Day ISO instructs the exchange to execute and/or post the order regardless of its view of away markets, and the post-only modifier instructs the exchange to reject or re-price the order if it would take liquidity before posting to the book. We suspect that HFT firms would find the post-only modifier particularly useful on IEX, as the odds that a non-marketable order becomes marketable en route to the exchange are higher given the extra latency created by the speed bump.

It also appears as though IEX intends to retain TOPS as its only proprietary data feed. TOPS only shows aggregated top-of-book order information, which would make it difficult for HFT firms to know where their orders are positioned in IEX's displayed order queue. Knowing where their orders are positioned in the queue helps these firms manage their risk.

For all these reasons, we don't expect to see much HFT liquidity in IEX's displayed pool, particularly not when these firms can provide liquidity more safely by using the hidden peg order. We think they'll continue to make displayed markets and pursue arbitrage opportunities on the incumbent exchanges,

while ignoring IEX as a displayed market and treating its dark liquidity as a niche trading opportunity much like they do today.

THE DISPLAYED POOL: AGENCY BROKERS MAY FIND POSTING ATTRACTIVE

Agency brokers could be well served by posting displayed child orders on IEX when working the parent orders passively. They're accustomed to latency, so the speed bump won't necessarily scare them off. And then there are a host of potential benefits:

1. Their orders would have priority over equally-priced hidden orders.
2. If displayed liquidity from HFT firms is scant, as we expect, their orders would rest higher up in the displayed order queue than on a typical exchange. All else equal, being higher up in the queue increases the odds of capturing spread.
3. It sounds as though IEX is planning on discounting the trading fees for posting and consuming displayed liquidity (they hint at this in their first and second comment letters).
4. The rumored low take fee might attract some incremental marketable flow.
5. Only displayed orders will be capable of attracting child orders from the IEX router under the new order-handling procedures for routable orders.

Against these benefits, agency brokers would have to weigh the risk of executing at soon-to-be-stale prices. But really, that risk exists on all the exchanges, and it's probably more acute on high-rebate/high-take fee exchanges. Meanwhile, as we hinted earlier in the report, "trading at stale prices" is a short-term alpha concept that isn't always relevant for an agency broker that actually needs to trade so that its investment-manager client can build or liquidate a position. Posting displayed on IEX is a surprisingly intriguing idea for agency brokers, though the value of such a strategy will diminish to the extent the new routing scheme curtails the stream of marketable flow into the pool.

Clearly there are a lot of moving parts to the proposed IEX exchange. Predicting just how they might fit together is beyond our powers of foresight, although we hope our observations and speculations have provided some food for thought. We won't really know what type of liquidity ecosystem IEX is capable of generating until it goes live as an exchange.

PART II: THE IMPACT

Until now, we've focused solely on IEX: who benefits from it today, and who might benefit from it in the future. In this section, we consider the broader impact of IEX exchangehood on market quality and the overall US equity trading experience.

First, some perspective: porting IEX's less-than-2% market share from the ATS realm into the exchange realm is not going to change much. Even if it gains market share rapidly, we'll still be looking at an NMS that is roughly two-thirds exchanges, one-sixth ATSs and one-sixth everything else – give or take a few percentage points. It will continue to hum as a fast, fragmented, algo-dominated system.

What we're looking for are fissures in the dam: small weaknesses that could grow bigger over time. To that end, we see two issues with IEX gaining full-fledged exchange status that could potentially be problematic:

1. Granting IEX's quotes trade-through protection could prevent the NMS from becoming more efficient in the future
2. IEX's model could set a precedent of exchanges interfering with brokers' pursuit of best execution

While we focus on these negatives in the remainder of the report, it's only fair that we salute IEX for its entrepreneurial spirit, its innovative model that delivers real value to the marketplace, and its commitment to improving market structure.

TRADE-THROUGH PROTECTION: MORE COSTS THAN BENEFITS

Whether motivated by the threat to their own business models, or by a genuine desire to preserve the integrity of the US stock market – or both – exchanges and trading firms have roundly criticized IEX's attempt to bring its model into the exchange realm. The ensuing back-and-forth between IEX's supporters and its detractors has shined a light on aspects of IEX's model that appear to violate existing rules and conventions. The most controversial of these is whether IEX's displayed quotes and orders should be afforded trade-through protection.

IEX's detractors claim that the Commission is opposed to giving trade-through protection to exchanges that have deliberately introduced latency into their order-handling processes. As proof of their claim, they point to a single sentence in the 523-page Reg NMS adopting release:

> The term "immediate" precludes any coding of automated systems or other type of intentional device that would delay the action taken with respect to a quotation.

If IEX's speed bump is deemed to be such an "intentional device," it would not be acting "immediately" to execute, cancel or acknowledge quotes and orders. If its reactions to quotes and orders are not immediate, its own displayed quotes should be "manual" and not "automated" per Reg NMS Rules 600(b)(3) and 600(b)(37). Manual quotes are not afforded trade-through protection, which means that brokers wouldn't have to connect to or trade on the IEX exchange (although a desire to achieve best execution might oblige brokers to connect).

IEX has tried valiantly in its comment letters to argue that the speed bump is not actually an intentional device, but that looks like wishful thinking to us. The problem is that the Commission makes it clear that anything can be an "intentional device" that is in fact an intentional device and that delays action taken with respect to a quotation – and it seems to us that the speed bump meets these qualifications.

IEX also argues that if the speed bump is an intentional device, then so are the incumbent exchanges' points of presence, and so are the delay coils they use to equalize latency among their co-location customers. All three of these features, IEX contends, amount to prescribed physical distance between the customers' servers and the matching engine. Thus, if all three features are essentially the same, and one of them (the speed bump) is deemed an intentional device, then they should all be so deemed.

These points also seem wide of the mark. NYSE and Nasdaq and Bats don't force brokers to connect via their points of presence as IEX does, which means those exchanges don't force brokers to experience latency as IEX does. Any broker willing to pay for co-location services can ignore NYSE's or Nasdaq's points of presence completely. Bats doesn't offer co-location services, which means that brokers can

park their servers next to the Bats matching engine in Secaucus without paying Bats anything beyond the $400/month port fee that all customers must pay. The fact that Bats offers a point of presence in Weehawken for the convenience of brokers whose servers are located there does not mean that it has built latency into its system – not in the way that IEX has at any rate. More to the point, nothing about the distance between a point of presence and its corresponding matching engine looks remotely like an intentional device.

Going back to Bats, it's worth noting that although Bats itself doesn't offer co-location services, Equinix, which owns the NY5 data center in which the Bats matching engine is located, does. And Equinix doesn't offer it up for free. We understand that Bats negotiated to have Equinix offer co-location services to Bats' customers at a discounted rate, but this only underscores the value of the space immediately surrounding exchange matching engines. Bats decided that it didn't want to be in the business of leasing out that space; NYSE and Nasdaq decided that they did want to be in it. Either way, somebody owns and is leasing out the space at rates that exceed the going rates elsewhere.

In other words, co-location costs more because the space is inherently worth more. IEX ingeniously uses the latency of the speed bump to devalue the space around its point of presence (and its matching engine), and this is what levels the playing field. This is how IEX achieves fairness. But that's only true for its hidden orders. For its displayed orders, the space around its point of presence still has value because getting order messages there with haste can improve outcomes (see the above section on IEX's displayed pool not being friendly to HFT firms for more on this). In this regard, IEX is simply going the Bats route and letting Equinix profit by leasing the space out.

Incidentally, if the SEC wanted to level the playing field without IEX's ingenious but latency-adding innovation, it could do so by strictly regulating all connectivity charges. This would include lease rates for server space in all locations. That kind of incursion into a highly developed corner of the private sector would have ramifications, of course, and we suspect that the SEC doesn't want to go there. But it's an option.

As for the delay coils that the incumbent exchanges use to equalize latency among their co-location customers, it's true that they're also intentional devices that delay action taken with respect to quotations. The SEC would be justified in revoking the incumbents' protected-quote status for the same reason that it would be justified in denying that status to IEX. We don't think the SEC would do that, however. The magnitude of the latency introduced by the delay coils is tiny – single-digit microseconds – and therefore we highly doubt it impacts any firm's trading experience or routing decisions. Furthermore, the coils actually make it possible for the exchanges (and data centers) to offer co-location services to all comers on equal terms. Without them, the brokers with servers closest to the matching engine would have an advantage over the brokers with servers on the outer edges of the data center. And the exchanges could conceivably charge them different rates. What is now a two-tier market would explode into a ten- or twenty-tier market.

To summarize, we think IEX's speed bump is an intentional device, and we think it is uniquely an intentional device. That said, it doesn't interfere with the *accessibility* of IEX's displayed orders. As we explained above in the section on IEX's displayed pool not being HFT-friendly, traders racing their orders to the IEX point of presence get access to the displayed liquidity that exists at the moment those orders

arrive – even though it takes another 350 microseconds for the event (trade, cancel, modification etc.) to actually materialize.

The speed bump only interferes with the timing of the event and the subsequent dissemination of acknowledgment messages and market-data updates. IEX *will* disseminate market-data updates to the SIPs without delay, but of course the SIPs incur a separate latency when consolidating and redistributing the data in their feeds. The regulator may treat this direct-to-SIP dissemination as IEX fulfilling its obligation to "ensure promptness of reporting" per Rule 601 of Reg NMS, but it's of little practical significance to the latency-sensitive trading community that glues the markets together most tightly.

If the SEC's stated opposition to "intentional devices" stemmed only from a concern over the accessibility of a displayed quote or order, we could see it ignoring the fact that the speed bump is an intentional device and granting IEX's quotes trade-through protection. However, if any part of its opposition stemmed from a concern over the "Reg NMS adherence" problems caused by stale market data – an increase in locked and crossed markets, pegged orders on other venues trading at stale prices, gaming, etc. – then the Commission would have grounds for denying that protection.

If the Commission did deny IEX's quotes protected status out of concern over Reg NMS adherence problems, it would be an awkward admission that the SIP feeds are superfluous – at least when it comes to helping market participants and venues avoid Reg NMS adherence problems. (If the SIP feeds were useful in this regard, then the fact that IEX is pledging to send its market data to the SIPs directly – not via the speed bump – would be enough to assuage any concerns.) Most market participants already know that the SIP feeds aren't useful in this way – it's an open secret – but it might be awkward for the Commission to have to acknowledge the fact.

On the other hand, if the Commission grants IEX's quotes protected status, it will have to contend with copycat rule filings from any incumbent exchanges wanting to play the speed bump game. Complexity would be a real concern, and the Commission would need to contemplate guiding the exchanges to acceptable types and magnitudes of latency. Otherwise, exchanges could attempt to introduce speed bumps that cater to certain segments of the market in a way that harms market quality.

In its Reg NMS re-proposal in December 2004, the Commission was clear that it didn't think guiding exchanges to acceptable magnitudes of latency was the right course of action (from section II(A)(2)(a) here):

> "The definition of automated quotation does not set forth a specific time standard for responding to an incoming order. The Commission agrees with commenters that the standard should simply be "immediate" – i.e., a trading center's systems should provide the fastest response possible without any programmed delay."

That's not to say that the SEC couldn't chart a new course, with the rationale being that the market has changed considerably since those words were written. But 2004 wasn't *that* long ago. It will want to consider whether any wisdom still lingers in its initial decision. We suspect that the primary concern, then as now, is that capping the allowable latency at a certain level would create a latency "floor" below which exchanges would lose competitiveness. For this to happen, the market share and revenues gained

by building a bigger pool of speed-bump-protected hidden liquidity would have to outstrip the share and revenues lost by having a less-competitive displayed quote.

We're not convinced that all exchanges would go for such a trade-off, but to the extent they did, latency would proliferate and price discovery would become less efficient. It goes back to HFT firms not coping well with the uncertainty over the status of their outstanding orders. If more exchanges implemented speed bumps, the uncertainty would become harder and harder to avoid, and eventually bid-ask spreads could widen out. A lot of dominoes would have to fall for this to happen, but that is the risk.

The SEC could decide not to give explicit guidance around latency types and magnitudes, and consider each copycat filing separately. But by accepting speed bumps it liked and rejecting those it didn't, it would simply be providing implicit guidance instead.

With or without explicit guidance, a decision to give IEX's quotes protected status looks to us like a recipe for more complexity and more Reg NMS adherence problems, and potentially more latency and less-efficient price discovery. The irony of the situation is that no one knows if IEX's displayed orders will add any value. The speed bump-protected hidden pool is the component that we know is innovative and that levels the playing field between those who can and can't afford co-location. The displayed orders will be uncompetitive almost by definition (because of the latency), so it's not clear to us what benefits, if any, would materialize if the SEC were to give them trade-through protection.

Meanwhile, being denied trade-through protection would come with a nice silver lining: rival exchanges would be less likely to roll out speed bumps of their own, because doing so would mean sacrificing their protected-quote status. That would leave IEX to dominate and perhaps expand the niche it's already carved out for itself.

COMPLICATING BEST EXECUTION

Brokers are in the business of accepting orders from clients and executing those orders in a way that helps the client meet its trading objectives. This process may include crossing stock with another institution, or executing in dark pools or on exchanges. Brokers use their knowledge of the execution options available to them, and their understanding of the client's objectives, to decide where to send the client's orders.

Exchanges, on the other hand, are public markets that exist mainly to bring together buyers and sellers. Historically, they have done so without judging or anticipating the value of the trade to the buyer or the seller. The "no judgment" aspect helps ensure that a market is always available to parties who aren't looked upon favorably by a particular broker, but who nonetheless have a legitimate trading interest. Technology and regulation have evolved in such a way that broker-dealers can easily create quasi-exchanges (ATSs), while exchanges can and do perform broker-dealer tasks like order-routing. But regulators have always made sure that where the lines between brokers and exchanges blur, they don't disappear entirely.

Our concern is that as IEX transitions from broker-dealer to exchange, it will bring with it broker-like functionality that all exchanges will then be able to mimic. All brokers would then be disadvantaged relative to exchanges, which could use their inherent advantages to compel brokers to outsource those broker-like functions to them. IEX's old routing scheme was the clearest example of broker-like activity

that might have been irresistible when combined with other elements of the exchange's offering. That advantage is no longer in play, but our concerns haven't been completely assuaged.

For not-held orders, brokers have the freedom to route where they please, when they please, subject to their duty to act in accordance with the principles of best execution. If they want to outsource certain functions that they're ultimately responsible for, like keeping a buy order pegged to the bid side of the market, they can use an exchange order type to do so. But the responsibility of executing at a good or fair price ultimately lies with the broker, not with the exchange. The same goes for an exchange's router: using one does not absolve the broker of its own best execution obligation.

If a broker wants to locate its servers in Secaucus next to the Bats matching engines, because it feels that that location provides the best optimization of cost, speed and centrality for its purposes, it can do so. If instead it wants to locate them in farm country in the middle of Pennsylvania because the space there is much cheaper and the latency doesn't impede its pursuit of best execution, it can do so. If it wants to co-locate in all three of the major data centers because that's what it feels is needed to achieve best execution, it can do so.

The mix of freedom and responsibility assumed by the broker works well because clients have different trading objectives. Some objectives require latency-sensitive strategies, and some don't. Some are price-sensitive, and some are time-sensitive. Some require natural liquidity, and some can be achieved by trading solely with intermediaries. Clients gravitate toward the brokers that serve them best.

Giving the broker wide latitude to pursue best execution may not always lead to the best outcomes, but we feel that it works better than the proposed alternative: exchanges that have no knowledge of the end-client's trading objectives manipulating the broker's message traffic, and preventing the broker from pursuing executions on away markets, in order to produce outcomes that they think will be good for the end-client. And of course the question looms large of how an exchange can "protect" one broker's client without harming another broker's client when those two end-clients would otherwise have traded with one another, but were prevented from doing so.

We want to be clear that this is not an attack on IEX itself. Our concern is that *in exchange form*, and as a precedent-setter for other exchanges, it threatens to limit brokers' freedom to pursue best execution as only they can pursue it.

CONCLUSION

At its core, IEX is a trading venue that segments out late-arriving marketable order flow, and it does this by holding up brokers' message traffic. It's ingenious, it benefits those who trade on it, and we value it as an ATS – but it shouldn't be an exchange with a protected quote.

If it were approved as an exchange with an uprotected ("manual") quote, it would take some getting used to, but we think that it might actually be a solution that works for everyone. IEX would gain Self-Regulatory Organization status and could take an active role in NMS governance. Its clearing costs would go to zero and it would gain immunity from liability when carrying out SRO functions. Most importantly, its business model wouldn't have to change at all – it would still protect investors' orders, and those of HFT market makers, in the same way that it does today.

The downside for IEX is that it would forgo any claim on the quoting share of the tape revenues, which accrues to SROs with protected quotes only. And any hopes it may have had to attract non-marketable limit orders from retail brokers, or corporate listings, would probably be dashed.

Back on the plus side, however, rival exchanges would likely be dissuaded from proposing speed bumps of their own. That would mean less competition for IEX, but, more importantly, it would assuage concerns about complexity and meaningful latency seeping into the NMS. Reg NMS adherence problems related to IEX's stale market data would be less troublesome, and brokers would not be forced to subjugate their own interpretation of best execution to IEX's (but could still route to IEX when their interpretations were aligned).

Intriguingly, amid calls from many corners of the industry to do away with the order-protection rule, the SEC would be able to gather data on how often and under what circumstances IEX's unprotected quotes were accessed or ignored. While not perfectly reflective of a market with no order protection whatsoever, the data might still provide some insight into the pros and cons of taking such an action.

Approval with manual quotes makes a lot of sense to us. However, we recognize that many in the asset management community want to see IEX in full-throated competition with the incumbent exchanges despite the negative effect that it may have on market quality. And we should be clear that any negative effect would be marginal given that we're talking about microsecond timescales. It's probably more accurate to say that the latency introduced by IEX's speed bump could prevent HFT firms from wringing the last remaining drops of efficiency out of our fragmented market structure. Put differently, order messages traveling at the speed of light with no friction upon entry into the exchange matching engine would be the maximum-efficiency state that may not be achievable in a market dominated by speed bumps.

All else equal, maximum efficiency is desirable. But it requires profit-seeking activity on the part of HFT firms to make it happen, and the suspicion is that the biggest HFT firms are extracting rents from the system – not profits. Some asset managers – and brokers – hope that an IEX exchange with protected quotes can change this dynamic, and restore the institutional investor's pride of place at least to some degree, without unduly harming market quality. We're not convinced that the trade-off will play out that cleanly, but we certainly understand the desire to achieve a better, fairer market structure.

APPENDIX: HOW THE SPEED BUMP WORKS

The IEX ATS is often described as a "dark pool," and for the most part that's accurate: most of its liquidity is comprised of hidden orders. Some of these hidden orders are block-sized midpoint peg orders, making IEX effectively a block-crossing engine for institutions. But its average trade size of around 215 shares, per data from FINRA, suggests that there are plenty of smaller sized algorithmic orders in the pool as well.

IEX is also a "lit pool," as it accepts displayed orders. Any broker-dealer consuming IEX's proprietary data feed, TOPS, can see the aggregated size of its top-of-book orders in real-time. IEX began accepting displayed orders in late February 2015, and recent stats from IEX's website suggest that these orders account for 8-10% of all shares executed on the venue.

Finally, IEX allows brokers to use its smart-order-routing technology to send the residual pieces of orders not filled at the venue itself to the exchanges with the best prices. We refer to this technology throughout the report as "the router" or "IEX's router."

In isolation, IEX's dark pool, lit pool and router function much like other venues and routers. What we and other market participants find valuable – and controversial – is how IEX employs a speed bump to delay message traffic into and out of these components.

COILED OPTICAL FIBER THAT DELAYS MESSAGE TRAFFIC

IEX's speed bump is a coil of optical fiber through which order messages (including cancel and modify messages) must travel before reaching the IEX matching engine. Messages that IEX sends back to the customer, including messages sent to customers through its proprietary data feed, travel through the coil in the opposite direction. It's approximately 38 miles long when stretched out, but it fits into a compartment no bigger than a shoebox when coiled. It takes a message approximately 350 µs (microseconds) to traverse the coil.

The speed bump is physically located in a data center in Secaucus, New Jersey, just inside the perimeter of IEX's system architecture. The system architecture is comprised of all of the information technology systems that the exchange employs to accept, acknowledge, execute and route orders, to send out and receive market data, to do risk checks, and to manage all the other processes that a trading venue must manage. Customers connect their own servers to IEX's servers in Secaucus, and order messages that customers send in – buy, sell, cancel, modify, etc. – must traverse the speed bump on their way out of Secaucus en route to the matching engine a few miles away in Weehawken, New Jersey.

IEX's system architecture is unique in the sense that customers aren't allowed to connect via servers in the same data center as the matching engine – a practice known as co-location. They can only connect via its servers in Secaucus, which along with the speed bump are known as IEX's point of presence or POP. To illustrate the difference between a more traditional exchange architecture and IEX's speed bump-enabled, co-location-disabled architecture, we've created simplified representations of both in Figures 1 and 2.
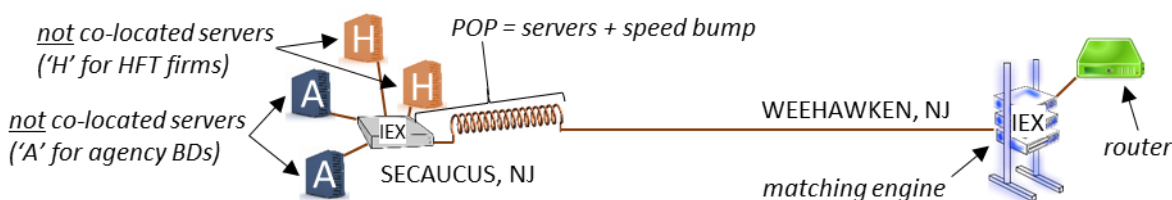
Figure 1.

IF IEX HAD A TRADITIONAL SYSTEM ARCHITECTURE (+BROKER-DEALER SERVERS)

Figure 2.

REPRESENTATION OF IEX'S ACTUAL SYSTEM ARCHITECTURE (+BROKER-DEALER SERVERS)

These schematics are of course only representative, and are certainly not to scale. That said, we believe that they convey an accurate understanding of the salient features of IEX's system architecture based on the information IEX has provided in its marketing material and in its comment letters.

ADDING REAL VALUE TO IEX'S HIDDEN PEG ORDERS

The speed bump is designed to prevent fast, opportunistic traders, whom we'll call "high-frequency traders" or "HFT firms" for simplicity, from trading against IEX orders that are pegged to an outdated or "stale" (or soon-to-be-"stale") National Best Bid or Offer. Trading at stale prices represents a kind of bait-and-switch for liquidity providers in the sense that the description of, for example, a "primary peg" order type leads them to believe that they'll be capturing the entire bid-ask spread, when in fact they may only capture part of it or in some cases none of it.

To understand how IEX protects hidden peg orders – and bear in mind that all IEX peg orders are hidden – consider an HFT firm that "sees" that the National Best Bid (NBB) in a given stock has just ticked down. If IEX had no speed bump, this firm might quickly route a sell order to IEX in the hopes of finding a hidden "primary peg" buy order to sell to at the old NBB before IEX itself "sees" that there's a new NBB. If such a buy order existed on IEX, and if the HFT firm got their sell order to IEX before the venue knew to lower the price of its buy order, the trade would go off at the price of the old NBB. The HFT firm would have established a short position (or closed out a long position) at a higher price, and IEX's hidden peg buyer would have paid more for the stock than they bargained for. The following diagrams illustrate how such a scenario, referred to as "latency arbitrage," might unfold if indeed IEX did not have its speed bump.

Figure 3.
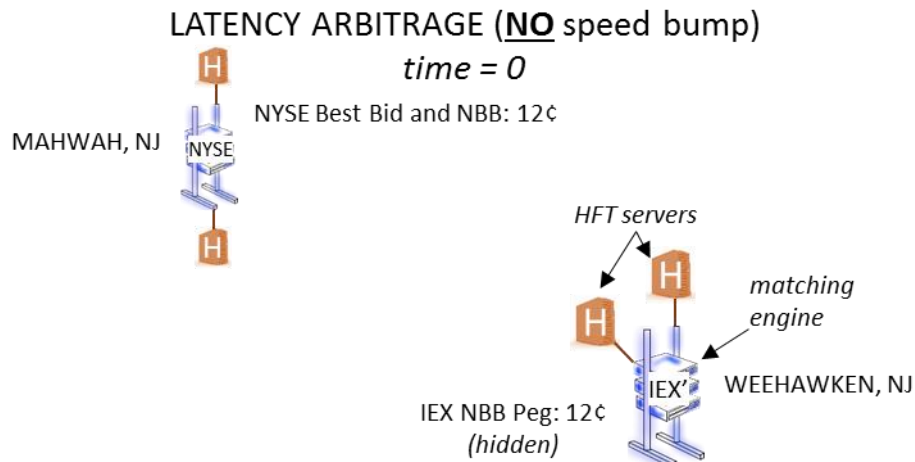
We've brought in NYSE's data center in Mahwah, New Jersey to create a hypothetical two-exchange market, and we've excluded the IEX POP and router as they're not relevant in this scenario. At time zero, NYSE's 12-cent bid is the NBB and IEX has a hidden buy order pegged to the NBB.

Figure 4.
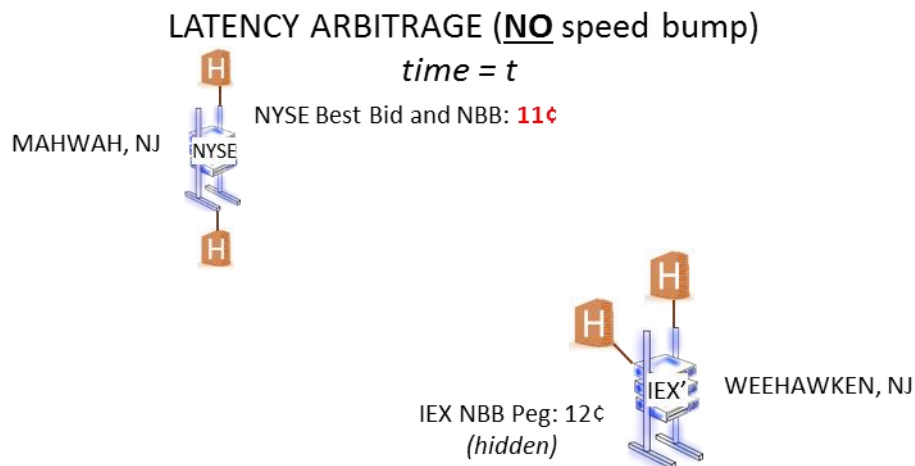
Suddenly, at time 't', a cancelation or trade on NYSE drops the NBB from 12 cents to 11 cents.

Figure 5.

LATENCY ARBITRAGE (**NO** speed bump)
time = t + 200 µs
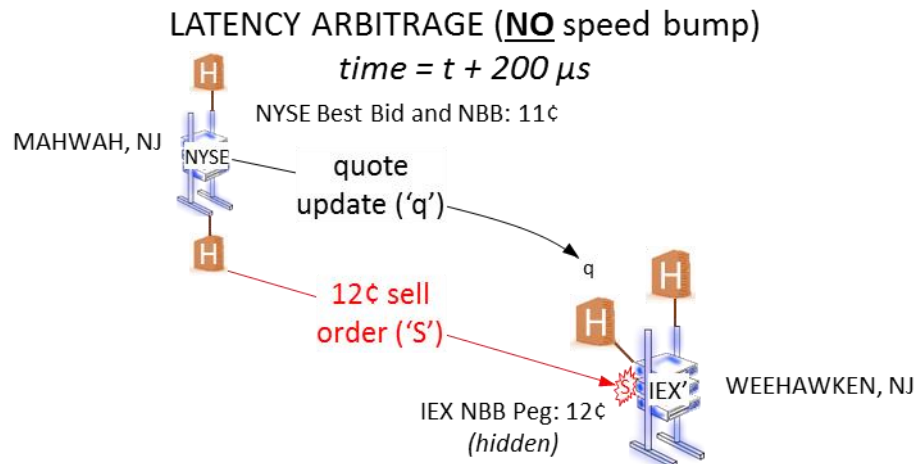
Almost instantaneously following the NBB change at time 't', an HFT firm routes a 12-cent sell order to IEX. Even though IEX is taking NYSE's proprietary data feed in this hypothetical example, the HFT firm's superior technology allows it to get its order to IEX's matching engine before the quote update message arrives. IEX "thinks" that the NBB is still 12 cents, so it executes the sell order against its hidden peg buy order at 12 cents. A fairer price from the buyer's perspective might have been 11 cents, the true NBB at the time of the execution.

Now let's take a look at how this scenario would have unfolded if IEX's speed bump was operational, which of course is the case today.

Figure 6.

LATENCY ARBITRAGE (**WITH** speed bump)
time = 0
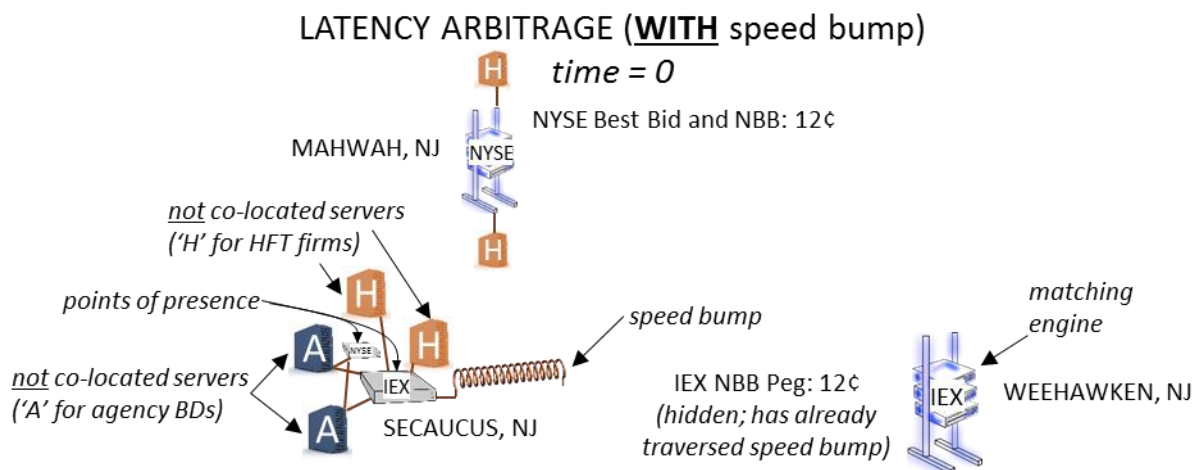
Notice that we've added points of presence in Secaucus for both IEX and NYSE, and that the HFT servers that clustered around the IEX matching engine in our earlier hypothetical now cluster around the IEX POP. At time zero, NYSE's 12-cent bid is the NBB and IEX has a hidden buy order pegged to the NBB.

Figure 7.

LATENCY ARBITRAGE (**WITH** speed bump)
time = t
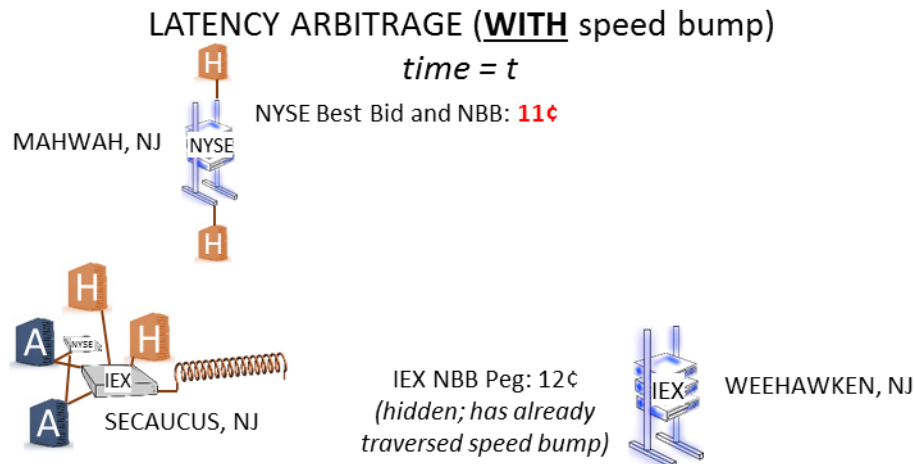
Suddenly, at time 't,' the NBB drops from 12 cents to 11 cents.

Figure 8.

LATENCY ARBITRAGE (**WITH** speed bump)
time = t + 250 µs
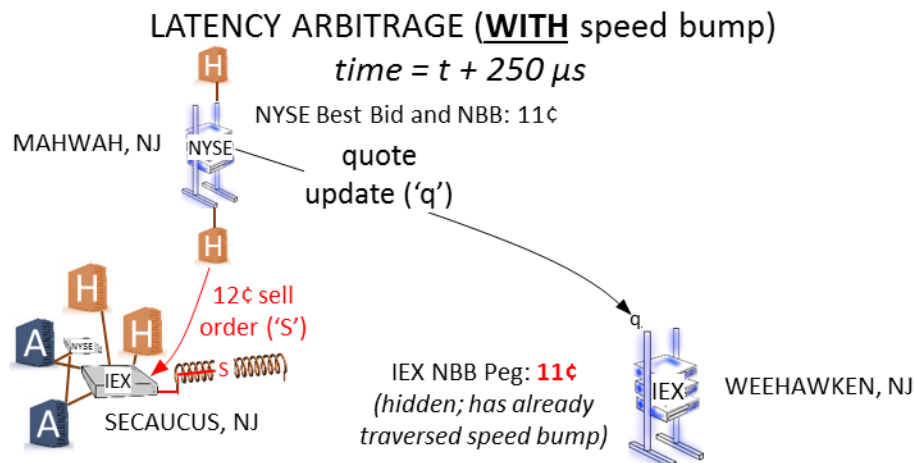
By forcing the HFT firm co-located at NYSE's data center in Mahwah to route its sell order to IEX via its speed bump in Secaucus, IEX ensures that the quote update coming directly from Mahwah arrives first. As shown by the red 'S,' the HFT firm's sell order is still spinning around the coiled optical fiber when IEX gets word from NYSE that the new NBB is 11 cents. In this scenario, because the HFT firm sent the order with a 12-cent limit, and IEX's best bid is now 11 cents, no trade will take place on IEX.

It's worth noting that the speed bump also offers protection against slightly different strains of latency arbitrage than the one we've diagrammed above. It prevents executions against hidden IEX orders pegged to stale NBBO *midpoint* prices, and it prevents a trader who suspects that a price is *about to become* stale from "picking off" a hidden order on IEX that is pegged to that price. The latter scenario is probably better described as aggressive or opportunistic trading, but it's still referred to by many – somewhat sloppily in our opinion – as "latency arbitrage."

ADDING REAL VALUE TO IEX'S ROUTING CAPABILITIES

The IEX router is effectively co-located in its Weehawken, New Jersey data center as depicted by the green server icon in Figure 2 above. Any residual piece of a marketable, routable order that isn't filled against IEX's own orders is sent without delay to the router with instructions to route the shares out to the exchanges with the best prices.

IEX doesn't publicly disclose the methodology by which it routes those residual shares, but given the router's high fill rates and IEX executives' well-documented knowledge of the latencies that exist between the exchanges' system architectures, we can safely assume that the basic idea is to schedule the individual routes so that each arrives at its destination at around the same time as all the others. Doing so would ensure that liquidity providers at one venue can't pull or change the price of their orders in response to executions taking place at another venue – a practice known as "quote fading."

We don't need to know the exact methodology of the router to know that the speed bump makes it more effective. While the above examples showed how the speed bump delays *inbound* message traffic, it delays *outbound* message traffic as well. The outbound delay prevents all market participants and all other venues who consume IEX's TOPS feed from knowing that an execution has just taken place at IEX. Because the traders posting liquidity at other markets have no chance to pull or adjust their quotes in response to news of the execution (i.e. no ability to "fade"), the IEX router has no need to race its orders out to these markets. It can take its time, knowing that the liquidity that it sees now is in all likelihood the liquidity that will be there a few hundred microseconds from now.
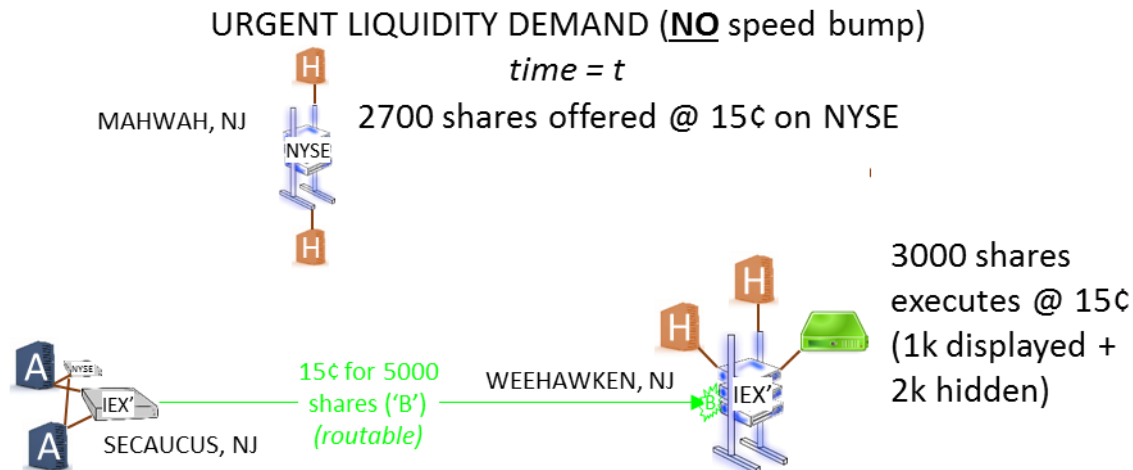
To illustrate how the speed bump empowers the router, we consider a scenario in which a trader has an urgent need to purchase stock and leverages the IEX router to do so. Figures 9, 10 and 11 show what would happen if IEX chose *not* to employ its speed bump, and Figures 12, 13 and 14 show what would happen if IEX employed its speed bump as it does today.
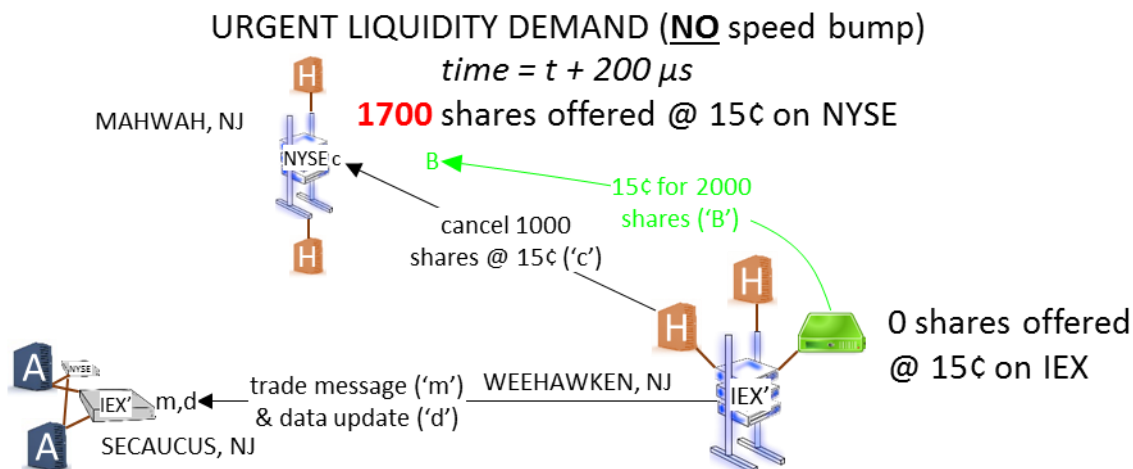
Figure 9.
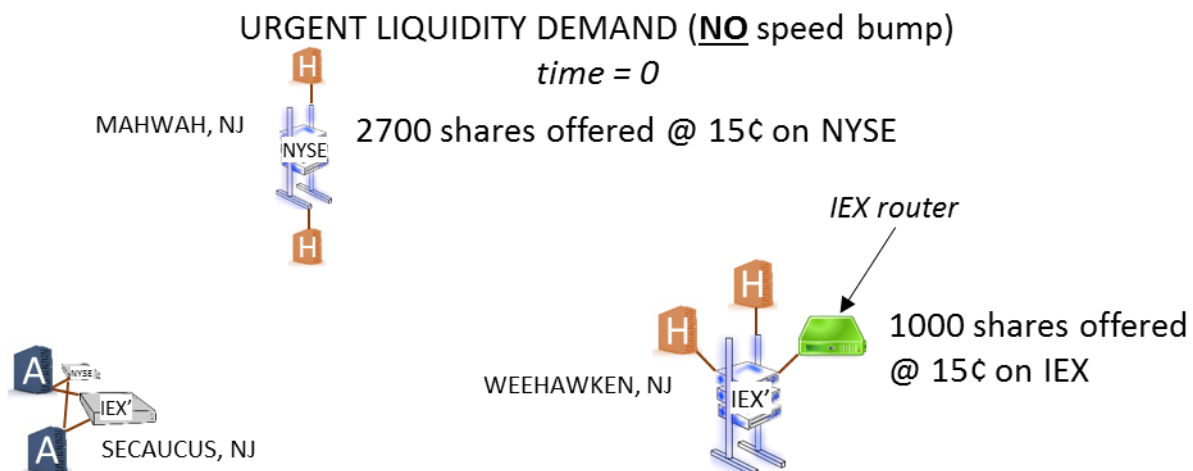
Note that we've added the IEX router to the schematic (see green server icon). At time zero, both IEX and NYSE are offering stock at 15 cents.

Figure 10.

URGENT LIQUIDITY DEMAND (**NO** speed bump)
time = t
2700 shares offered @ 15¢ on NYSE

MAHWAH, NJ

3000 shares executes @ 15¢ (1k displayed + 2k hidden)

15¢ for 5000 shares ('B') (routable)

WEEHAWKEN, NJ

SECAUCUS, NJ

At time 't,' a routable 15-cent buy order for 5,000 shares arrives at the IEX matching engine and matches against 3,000 shares on the IEX book, leaving a 2,000-share residual. Note that IEX was only showing 1,000 shares, but had 2,000 shares in hidden liquidity behind it.
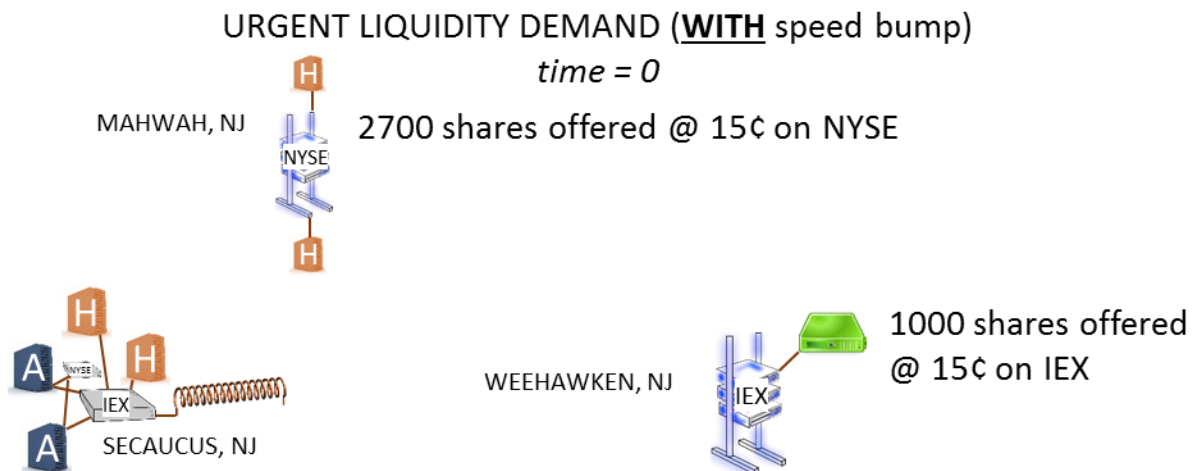
Figure 11.

URGENT LIQUIDITY DEMAND (**NO** speed bump)
time = t + 200 μs
1700 shares offered @ 15¢ on NYSE

MAHWAH, NJ

cancel 1000 shares @ 15¢ ('c')

15¢ for 2000 shares ('B')

0 shares offered @ 15¢ on IEX

trade message ('m') & data update ('d')   WEEHAWKEN, NJ

SECAUCUS, NJ

At time 't', IEX immediately sends a trade message to the buyer whose servers are in Secaucus, and also updates its market-data feed ('m' and 'd' in the schematic). At the same time, it sees that NYSE is offering 2,700 shares at 15 cents, so it instructs its router to route the 2,000-share residual there. Although the instruction is given almost instantly at time 't', the IEX router can't out-race the co-located HFT firm that is reacting to the execution by routing a cancelation message to NYSE. It might be that the HFT firm sold stock to the buyer in the 3,000-share transaction and doesn't want to risk selling more to that buyer on NYSE, or it could be that the HFT firm is simply unnerved by the presence of what seems like a large, aggressive buyer. In any event, by the time the IEX-routed buy order arrives at NYSE at time

't + 200 µs', the amount of displayed interest has dropped to 1,700 shares – not enough to fill the balance of the buyer's order.
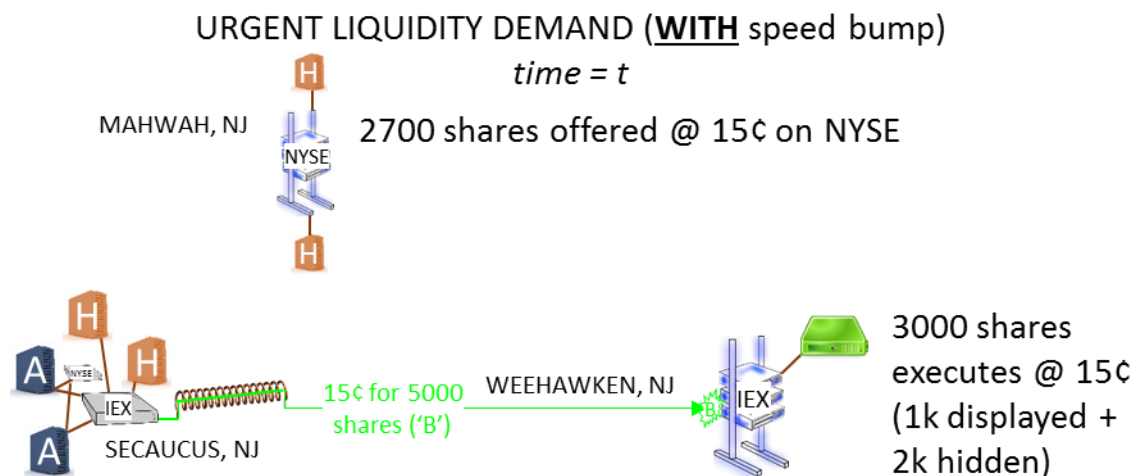
Now we consider how the same scenario would unfold if IEX were operating with its speed bump in place and without co-location. This is a representation of how IEX operates today.

Figure 12.

URGENT LIQUIDITY DEMAND (**WITH** speed bump)
time = 0
2700 shares offered @ 15¢ on NYSE
MAHWAH, NJ
1000 shares offered @ 15¢ on IEX
WEEHAWKEN, NJ
SECAUCUS, NJ

At time zero, both IEX and NYSE are offering stock at 15 cents. Note that the speed bump is now in place and that the HFT servers are now clustered around the IEX POP, which is the only point of entry for order messages into the IEX matching engine (and the only point of exit for proprietary market data and trade messages sent from Weehawken).

Figure 13.

URGENT LIQUIDITY DEMAND (**WITH** speed bump)
time = t
2700 shares offered @ 15¢ on NYSE
MAHWAH, NJ
15¢ for 5000 shares ('B')
WEEHAWKEN, NJ
3000 shares executes @ 15¢ (1k displayed + 2k hidden)
SECAUCUS, NJ

At time 't,' a routable 15-cent buy order for 5,000 shares arrives at IEX and matches against 3,000 shares on the IEX book, leaving a 2,000-share residual.

Figure 14.

URGENT LIQUIDITY DEMAND (**WITH** speed bump)
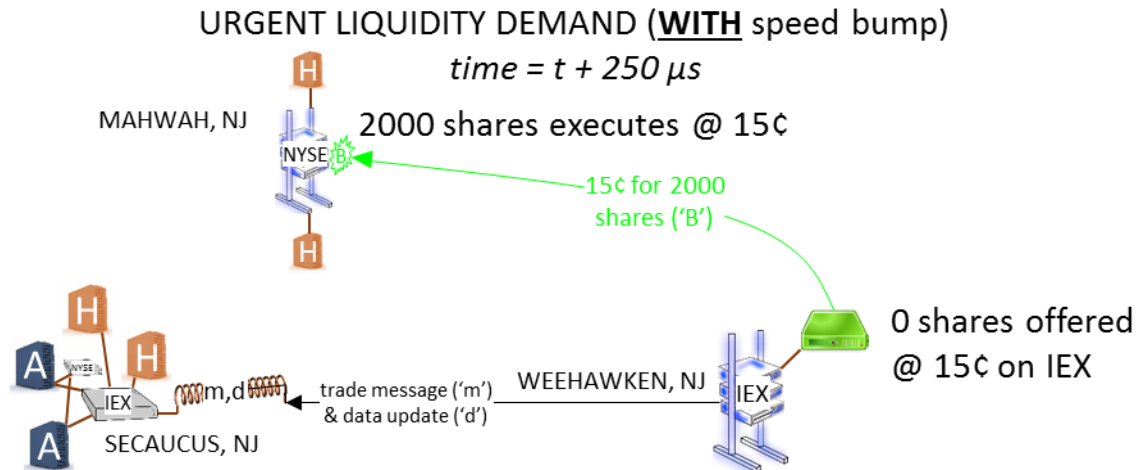
Still at time 't', IEX immediately sends trade messages to the buyer and seller whose servers are in Secaucus, and also updates its market-data feed. At the same time, it sees that NYSE is offering 2,700 shares at 15 cents, so it instructs its router to route the 2,000-share residual there. When the order is en route to NYSE, no one but IEX itself knows that an execution has just taken place on IEX. The trade messages and the proprietary data-feed updates are still traversing the speed bump en route to the servers in the POP ('m' and 'd' in the schematic), and the updates that IEX sends without delay to the Trade Reporting Facility (not shown in the above diagram) won't be processed in time to alert the HFT firms co-located in NYSE's Mahwah data center of the trade event. As a result, the buyer is able to buy all 2,000 shares at NYSE at time 't + 250 μs', completing its 5,000-share "parent order."

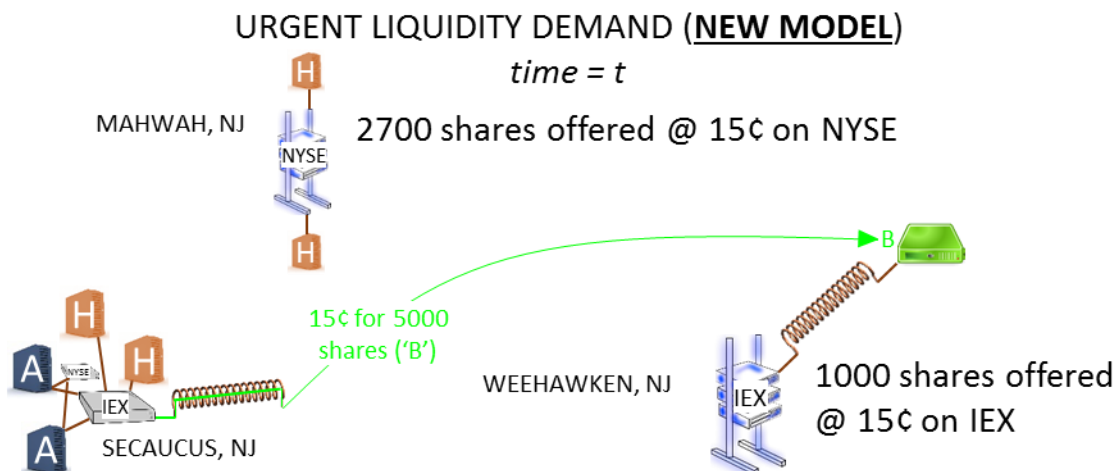THE NEW ROUTING SCHEME WILL LIMIT MARKET SHARE

On February 29, IEX announced that it will change how it handles routable orders once it becomes an exchange. Under the new scheme, once a routable order has traversed the speed bump, it will be sent directly to the router – not to the matching engine, which is where it is sent today. From there, the router will "look" at the market data it is receiving from all thirteen exchanges and decide on a schedule for the child orders it will send to IEX and away markets. In Figures 15-18 below, we illustrate how the new system architecture will work in the same 'urgent liquidity demand' scenario used in Figures 9-14.

Figure 15.

URGENT LIQUIDITY DEMAND (**NEW MODEL**)
time = 0
2700 shares offered @ 15¢ on NYSE
1000 shares offered @ 15¢ on IEX
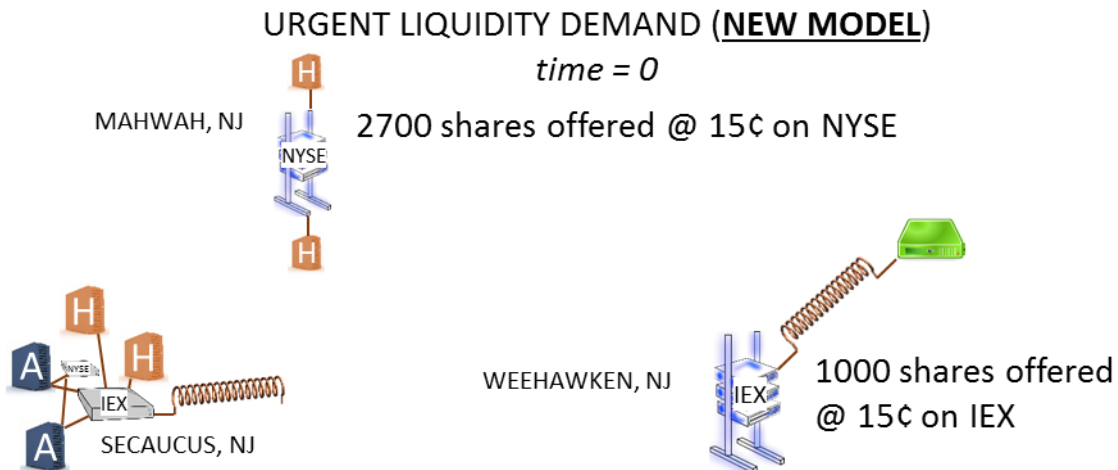
As before, we have a simplified two-exchange market with 1,000 displayed shares offered at 15 cents on IEX and 2,700 displayed shares offered at the same price on NYSE. Note that there is now a second speed bump that sits between the IEX matching engine and the IEX router.

Figure 16.

URGENT LIQUIDITY DEMAND (**NEW MODEL**)
time = t
2700 shares offered @ 15¢ on NYSE
15¢ for 5000 shares ('B')
1000 shares offered @ 15¢ on IEX

At time 't,' a 5,000-share buy order with a 15-cent limit arrives at the IEX router. The IEX router "sees" the shares offered on IEX and at NYSE. Knowing that it will take about 350 μs for a child order to arrive at the IEX matching engine and about 250 μs for a child order to arrive at the NYSE matching engine, it gives the IEX child order a head start.

Figure 17.

URGENT LIQUIDITY DEMAND (**NEW MODEL**)
time = t + 100 μs
2700 shares offered @ 15¢ on NYSE
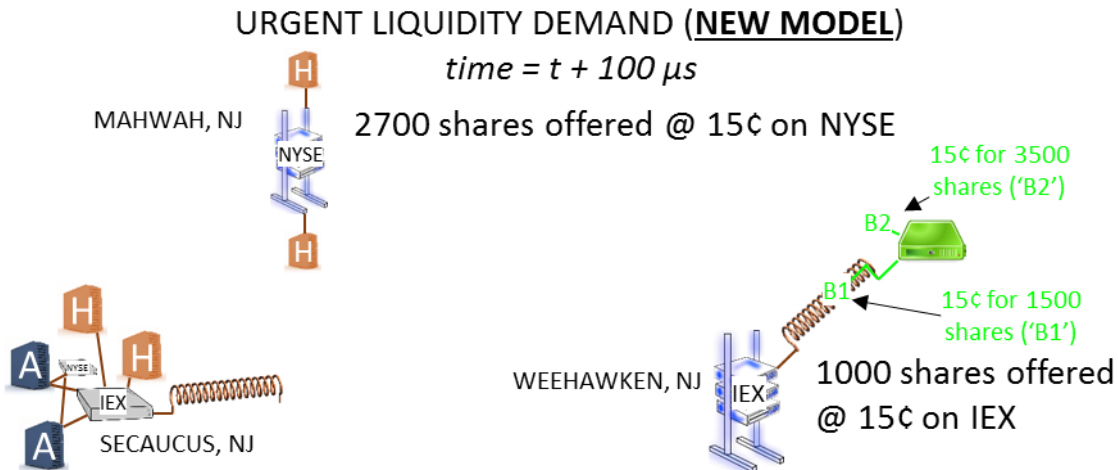
At time 't + 100 μs,' a child order is en route to the IEX matching engine via the speed bump (this is the green 'B1'), and the router releases the NYSE-bound child order (B2). Because the router only "sees" 3,700 displayed shares on NYSE and IEX, it has discretion over how to route the 1,300 shares that aren't spoken for. We assume that IEX will oversize both child orders – 500 shares extra for the IEX child order and 800 shares extra for the NYSE child order – in the hopes of finding hidden liquidity.

Figure 18.

URGENT LIQUIDITY DEMAND (**NEW MODEL**)
time = t + 350 μs
2700 shares executes @ 15¢
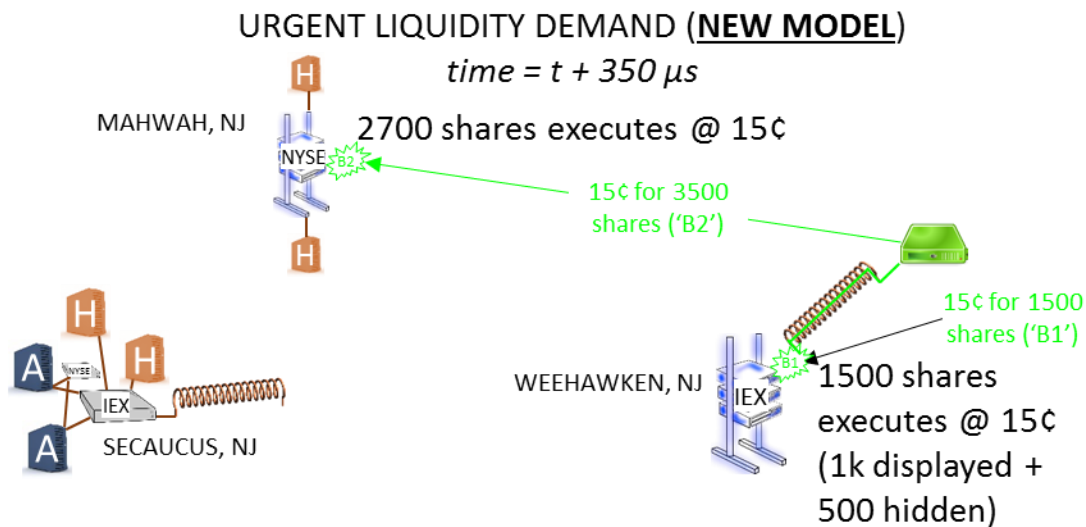
At time 't + 350 μs,' both child orders arrive at their destinations. The buyer has purchased only 4,200 shares – not enough to complete its 5,000-share order – because hidden liquidity was found at IEX but not at NYSE. At 't + 700 μs,' the buyer will receive news of the execution on IEX, and 250 μs later, at 't + 950 μs,' the buyer will receive news of the execution on NYSE by way of IEX (these messages aren't shown in the diagram). Since it took 350 μs to get the order to the router in the first place, users of IEX's

routable orders can expect to wait a total of 1,050 μs for news of any executions on IEX. That compares to the 700-μs wait that users of routable orders must endure today, and that users of *non*-routable orders must endure today and in the future. The wait for news of executions at away markets is a bit longer with the new model too, since the router must hold up any child orders sent to those markets to ensure that they arrive at around the same time as any child orders sent to the IEX matching engine (assuming the router does send a child order to IEX).