

From the SEC act of 1934

**It shall be unlawful for any person, directly or indirectly, by the use of any means... ..To use or employ, in connection with the purchase or sale of any security registered on a national securities exchange ... any manipulative or deceptive device or contrivance in contravention of such rules and regulations as the Commission may prescribe as necessary or appropriate in the public interest or for the protection of investors.**

HFT ( sub second ) problems and solutions:

SEC, here is **one very simple rule** to implement and enforce that will make predatory sub second HFT problem disappear and will automatically become very simple "**speed limit**" rule, will build a bridge between human and machine trading without slowing down executions. It's the proposed rule that a lot of traders have been talking about on trading boards and in private, basically it's very simple and makes perfect sense.

Let's call it **cancel/replace quote rule** has 2 parts:

after a quote ( an order ) reaches the exchange matching engine, the quote is held for 500ms before it can be canceled/replaced/modified ( part 1 ), but the quote can be executed against without any delays, not to slow down the execution ( part 2 ), basically 500ms delay before you can pull ( cancel ) or replace or modify your quote, this 500ms does not apply to executions, only to cancel/replace/modify.

Everybody is talking about the need to slow down trading to allow humans to compete with machines these days. Average human reaction time is 200ms, average fastest human reaction time is under 100ms, coupled with average internet latency worldwide of about 150 ms, we got about 500ms or half a second is needed for most traders:

1. for the quote to appear on traders/investors screen - delivery latency delay of 150ms and 100ms for traders average hardware and platform processing delay, times 2 for round turn
2. min of 200ms for trader/investor to react to the appeared quote, thus a total sum of 500ms ON AVERAGE and aggregated across all traders/investors, all quotes, all the time.

For hft bots the reaction time is under 30ms including delivery latency and decision reaction time! Thus humans can not compete with machines. A bridge is needed, more time to allow humans to react to machine trading, this is 500ms cancel/replace rule where 500ms will be just enough time for humans to react to any quote in aggregate. Basically it comes down to this, every quote is **exposed** for 500ms on the matching engine, unless it's executed against, thus executions are not affected, executions can be matched faster.

In general there are 3 ways you can slow down trading:

1. You can slow down quotation system to allow human traders to have a look at what the machines are trying to do with the quotes and observe the matching engine, basically to

allow human traders at least to observe the process of how machine works with quotes in general and gives human traders min time to react to those quotes.

2. You can slow down executions by slowing down the matching engine, that goes counter the notion of ideal execution is instantaneous execution, this make less sense but should be explored in second phase.

3. You can 1 and 2, which is not the goal here.

SEC should start with 1 and later after 1 is invoked to examine 2 as well.

The 500ms rule only applies to 1, and not to 2.

The impact of this proposed rule on sub second hft:

Predatory sub second hft works on the principle of fast quote pushing in and pulling out, they can manipulate the market impact with or without trading just by quote flooding/stuffing/moving/churning, they avoid detection by pulling quotes out fast enough so that slower hardware or traders can not detect them at all. With 500ms rule their quotes ( unless executed against ) will now be exposed to everybody for at least half a second before they either pull them ( cancel ) or replace ( update the order properties). This 500ms will allow average hardware and average trader/investor around the world to react to their quotes. Basically hft firms now will be exposed for at least half a second, they simply can not afford that and thus quote flooding, stuffing and manipulation will seize right away because of the 500ms exposure they will be detectable by average hardware and human traders/investors.

Enforcing the rule is trivial: make the rule part of a requirement for every matching engine on nasdaq and nyse including all ECNs, applied to everybody from market makers to hft firms to brokers to traders/investors, without exemptions! Because it's the part of the matching engine, the rule is universal and there is no enforcing it, except maybe varying the 500ms parameter, say from 300ms to 1 second, based on feedback from traders later, see below for details.

The rule does NOT slow down executions, if execution against the quote happens before 500ms then let it happen, it ensures no slowdown in execution flow, will guarantee no lag of equities behind futures. The rule will get rid of junk and toxic quote/order flow. All hft firms will think twice before getting exposed for 500ms out there. Human traders will be see and react to what hft machines are doing, that's the main goal here, for human to be able to observe and track machines, else we can't register what machines are doing and we can't react to them, thus we have no control over machines - that's danger right there !

There is still going to be competitions for faster executions, but with clean order flow, toxic order flow will be gone. This will turn hft firms away from predatory trading and towards honest competition for order flow and more efficient executions. This will also clean up the quote system bandwidth and junk quote data. The rule is self enforcing, simple to implement as a programming operation in the exchange matching engines.

**The goal of this rule is to EXPOSE any quote for minimum duration of half a second so that not only machines but human traders/investors can see it and react to it.**

The rule does not slow down quotation system, it's simply minimum 500ms exposure for not yet executed quotes, that's it. It will be most effective for quotes on best bid and offer. The rule should not apply to time of sales data, there should not be any delays in reporting trades back to traders.

The rule should not apply or impact the market orders, as those are be executed without any delays, zero delay, but the rule will apply to any type of resting/passive order.

In simple term, you put an order out, that order is exposed for 500ms before you can do cancel/replace/modify it, unless it's executed sooner than that. It goes towards the notion of, if you put an order out there then let everybody see it for at least half a second, don't play any games with the quote system, like push-pulling, flooding, stuffing, front running, everybody can see you now for half a second, make sure you understand what you are doing, can't play old games anymore, people will see you and they take you out if they need your size.

This will not kill off collocation, there is no danger in that. Collocation will still be necessary to minimize the delivery latency and the closest to the exchange will still have the execution edge over those remotely located. Collocation will get your order in faster and thus can get you executed faster and get you data faster, but you are still going to be exposed for 500ms for everybody to see unless you are executed against, even if you have zero latency you are there on the matching engine for 500ms and thus you are exposed for absolute minimum time for everybody around the world to see you, so invisibility is out of the question. HFT becomes then the competition for clean order flow, who can get to the exposed quote faster to get the execution but not to manipulate the order flow with useless quotes. Collocated hft firms will still have the execution edge but as collocation becomes widespread the edge will even out between the market participants.

The main goal of this 500ms rule not to slow down executions but to clean up the quotation system of toxic manipulative order flow. Later, SEC should study the 2nd method because even after the quote system is cleaned from toxic hft order flow, trading executions can become so fast that market can move considerably over sub second period of time. However cleaning up quote system with minimum time exposure rule will kill off majority of predatory quote manipulations and thus will bring larger traders back from the dark pools into the main market, increasing liquidity and stabilizing the order flow, even with fast executions due to depth of market at best bid offer stocks won't move as fast through the levels. But in general something like 1ms execution interval or around there or about 500-1000 times faster than the 500ms quote rule should be a good number. SEC should experiment with that as well, but initially there should not be an speed limit on executions, to prevent any execution delays, method 2 must be studied on some stocks to see the result before making any decisions. Slowing down executions means slowing

down matching engine, this goes counter the notion that near instantaneous executions are highly desirable. The 500ms exposure rule does not slow down the matching engine, order matching speed is not a function of quote exposure but a function matching engine speed, as a matter of fact cleaning up toxic order flow will probably speed up matching engine executions as it will free up processing time due to unclogging of the quotation system from the toxic quotes.

Most traders would not want to slow down executions! No trader would want to slow down quotes "on the way in"! No trader would want to slow down time of sales data on the way out! Most traders/investors would want some kind of minimum exposure time before you can pull your order out or replace/update the order, because it's a natural thing: if you place an order - your order is on the exchange a minimum of half a second before you can pull/replace it, at same time your order can get executed without any delays and 500ms does not apply to executions, this gives everybody minimum time to see your quote ( machine or human generated ) and act on it, else don't mess with trying to get in and out 10 thousand times per second trying to manipulate the order flow, get in, get your execution or wait for it, if you made mistake with the order, wait 500ms and pull your order out and don't try to make mistakes next time and if your order is not executed within 500ms and market runs away and your order is left behind then first get your execution with a new order and then after 500ms passed you can pull the old order out. If the market is jerky is you got execution on the new order as well as the old order and you got twice the size you wanted then it's just tough luck, that's just the market, it happens sometimes, unload your half and go from there.

Counter argument against this rule:

some traders will say that this will expose their quotes for whole half a second and collocated hft firms will now have more time to decide what to do with the order, that's true. But let's not forget that hft orders will also be exposed for half a second as well and if they decide not to take the other side of the order and not execute against it. Everybody gets exposed for everybody else to see for half a second. If hft firm decides to take my exposed order out, more power to them, I get faster executions that way, let them compete over who gets first to my order, I win as I get faster execution against my order. If they see my order first and decide to front run me then fine, my order will be left behind and I will change or pull it out 500ms later, but they will get exposed doing that for half a second as well for everybody to take them out or to front run them in turn. As long as everybody gets equal exposure time there is no problem there, it equalizes. It becomes true hunt for liquidity and not an empty con game of manipulating thousands of quotes per second.

This rule will most likely increase the efficiency of executions, as it will unclog the quotation system and remove the toxic useless order flow. This may temporarily short term increase the micro volatility in the order flow after all predatory and junk hfts will leave and until firms adapt to this rule, slowly size will return back to main markets from dark pools, liquidity will deepen and order flow will normalize and micro volatility will subside. Some sort of transitional period is needed where the markets should be watched closely. SEC may enable the rule gradually starting from the least liquid stocks where

there are more problems observed and then to the most liquid and widely traded where there are less problems.

The rule can be controlled by the exchanges through one parameter  $De$  = quote exposure delay = average 500ms, if  $De$  is set to 0 then the rule automatically does not apply, thus rule can be enabled disabled by switching parameter to 0 or away from 0 where rule can be adjusted based on trader feedback. Optimal  $De$  value should be about 300-1000ms or from half a second to 1 second, but not more than 1 second, and not less than 300ms as traders should be able to pull/replace their orders after a second and not wait longer than that, while less than 300ms exposure won't be effective as most traders won't have time to react in aggregate. Typical formula for optimal value of  $De$  should be =  $2 \times$  average internet and average hardware processing latency + average human reaction time. Currently  $De = 2 \times (150\text{ms} + 100\text{ms}) + 200\text{ms}$  or about 700ms ~ roughly half a second will do.  $De = 1$  unit of average global internet latency delay of 150ms to get the quote delivered from matching engine to the average trader's average machine and 100ms for machine processing ( trader/investor + broker ) for the quote to be displayed on the average human traders screen given average hardware + 200ms for average aggregate human trader/investor to react in aggregate to that quote + 1 unit of average hardware processing latency 100ms + 150ms internet latency to deliver the traders hypothetical aggregate reaction to the exchange so that the quote is still there by the time the traders matching quote reaches the exchange. Numbers are in aggregate, for all hypothetical traders reaction times, over average traders hardware and over aggregate average global internet latency.

Average human reaction time stays the same, while average internet and hardware latency is gradually decreasing due to faster hardware and thus with time exposure delay can be adjusted to a lower value as hardware latency decreases with time. Intuitive values are from half a second to a full second, not less and not more. Theory goes that computer hardware that powers the internet is naturally designed in such a way that average internet round turn latency is naturally within the average human reaction time or < 200ms for real time interactions across the globe such as gaming or trading. More latency and unnatural delays will not enable real time interactions. Lower bound is zero but ideally less than 100ms delays are not detectable by human eye anyway, only by a machine, so naturally optimal human range is from 100+200 ( about  $2 \times 50\text{ms} + 200\text{ms}$  ) or around 300ms round trip minimum, ( this allows for 50ms one-way hardware processing and delivery latency delay ) less than that and humans won't care because they can't detect movements faster than that anyway, more than that and humans get irritated by delays in processing. So eventually within very short period of time processing and latency delay round trip will get below 100ms ( some may argue that it's already happening ) and humans will be content with that and won't care much below that. This means that min  $De$  is about 300ms while max  $De$  is about 1 second. Age of investor/trader must be considered as well as average human reaction time is a function of age as well, not only location and hardware latency. Basically we are talking  $De$  in the range of 300ms for younger human traders/investors with fast computer hardware and fast connections to about 700ms for older traders/investors, slower hardware and average internet connection, with 500ms median. Below 300ms humans won't be able to

compete with machines over quotes in aggregate, they won't see the changes in them in aggregate and won't be able to react to them in aggregate. Even time of the day is a factor in traders reaction times, towards the end of the day traders will typically get slower with reaction time > 300ms. Bottom line the best De range from 300 to 1000ms. Optimal value is around 500ms or half a second for aggregate traders based on age, time of day, location, connection and hardware. Safe value is about 700ms to 1 second. Over 1 second and human traders may get annoyed by delays in pulling/cancelling quotes out or changing them. Less than 300ms and even best human traders/investors with fast hardware and connections won't register any change or have time to react to each other's quotes in aggregate.

[http://ipnetwork.bgtmo.ip.att.net/pws/global\\_network\\_avgs.html](http://ipnetwork.bgtmo.ip.att.net/pws/global_network_avgs.html)  
<http://www.internetpulse.net/Main.aspx?Destination=Verizon>  
<http://www.internettrafficreport.com/>  
<http://www.verizonbusiness.com/about/network/latency/>

human reaction times:

<http://www.humanbenchmark.com/tests/reactiontime/stats.php>  
[http://en.wikipedia.org/wiki/1\\_E-1\\_s](http://en.wikipedia.org/wiki/1_E-1_s)  
[http://en.wikipedia.org/wiki/Reaction\\_time](http://en.wikipedia.org/wiki/Reaction_time)

To visualize the rule is simple:

imagine a future robot doing some work competing with a human, if robot is so fast that a human eye can not register what the robot is doing then human can not observe what robot is doing period, the human then can not control or compete with the robot as humans can not observe/react to what robot is doing, except to pull the plug ( circuit breaker analogy ) - that's dangerous. The solution is then to slow down the robot to the level that human eye can observe the robot operations, especially if robot is to compete with other humans or to participate in the important process of information factoring. This basically slows the robot down just enough so that humans can observe, interact, track, control, and compete with the robot. How do you slow down the robot? You make his movements slow enough for the human eye can register them. That's the analogy for the 500ms quote exposure rule. You expose the robots movements for minimum of 500ms for human eye to see and react to them.

Hft ( sub second trading scales only ) bots are simple mechanical trading devices that process simple order flow information on ultra short term time scales, unlike human traders hft bots can not process complex information inputs. They are not good at global information factoring ( economic, fundamental, sentiments, intra and inter market analysis ) and thus can not be trusted in pricing mechanisms. To put it plainly, hft bots are designed to be fast but not smart. Humans are slower but they are much smarter and can factor a lot more information than a machine can. Thus allowing bots to dominate trading is a dangerous game, they are fast and they are stupid and have no clues about limits or real information or people or volatility.

Flash crash of May 6th is a good example:

Market heads for massive profit taking after a huge run up from the bottom, as the selling wave starts to accelerate, price movement speeds up to the point that most human traders can not react to such trading speeds, what they observe ( if they observe it at all ) they can not react to, so they pull the plugs, they stop factoring information and stop trading, the only crowd dominating trading are stupid but fast trading bots then, trading then speeds up even more, then some slower trading bots ( operating intraday ) can not handle even that speed, humans pull plugs on them too, finally only sub second hft bots dominate the trading, and those start to break down, humans then issue commands to hft bots to stop trading, only super fast machines remain in the market, liquidity drops even more, trading speeds accelerates even more, until process runs out of bounds, which results either in no one left trading and a super crash where a few machines chase each other down out of control until the plug is pulled. Thus the process runs off and does not reflect economic reality and the only way to stop it to pull the plug or to run against the market close or with a circuit breaker.

Traders should be allowed to compete with machines, given enough time to observe what machines are doing and if necessary to step in to act without fear. Even if machines get as smart as human traders at factoring information which is unlikely ( as information cascades and multiplies over time and systems become more complex, human brain may still evolve faster than it can design the smartest hardware to factor that cascading and accumulating information ) , they should not be allowed to trade faster than humans can track them, and they should not be allowed to become smarter than humans in the first place or machines will take over for good and not necessarily for the benefits of the humans.

Now, hft firms are going to fight this rule with all they got, they will say it's no good, that it's not going to do anything, or that it's going to slow down executions, all not true, but these will be attempts to derail the rule. Market makers, broker/dealers are going to try to get exemption from the rule as usual, which will create more problems if exemption is granted by SEC, they will now have no exposure delay while the rest will, this will give them tremendous edge, by getting exemption to this rule they will subvert the whole process and make it worse. The exchanges should be neutral in this matter as this benefits everybody in the long term, however they may resist the changes necessary to implement rule because hft firms give them some business revenue, but SEC must force exchanges to think about long term benefits of traders/investors and think about preventing potential machine induced crashes. SEC should guard against such attempts to loophole around the rule, the rule must be universal, non negotiable and applied to everybody without exemptions across all equity markets, else SEC will be considered at failure to act in solving the problem or actually making it worse. Same simple rule for everybody!

The nice thing about the rule that it's implemented on the exchange level, not on an individual firm level, thus the rule is self enforcing and universal and there is nothing more for SEC to enforce except to make sure each exchange or ecn has that rule implemented. SEC should also work with foreign equity exchanges to get same rule

passed there as well to be coherent across the globe. Markets such as electronic Globex are not affected by predatory quote manipulations as those matching engines are clean, centralized, not fragmented, can not be manipulated, are fair and well designed and thus the rule should not apply to those exchanges, unless sub second hft predatory trading is detected there rule should be enabled there as well before they can damage those markets.

Slow down the machines to the level of humans, else face the consequences! One simple 500ms min exposure rule will make that happen and will maintain the order in quotation systems without compromising the execution speed. Two tiered trading fragmentation into sub second hft and not must go. A bridge must be build between humans and bots. The choice is either clean up sub second hft sector or face massive flash crashes, problems, upset traders/investors, no other options. No action taken means competition for latency will heat up, trading frequency will increase even more, instead of tens of thousand of useless quotes per second it will be hundreds of thousands, speed will increase even more and with that probability of flash crashes will rise with each millisecond decrease in latency.

SEC should do the study on the optimal quote exposure value, considering all factors: human reaction times, traders/investors average age, average traders hardware processing latency, average global internet latency, any other factors that might have been missed here.

HFT with times scales of over 1 second ( typically intraday trading on time scales from 1 second to full day ) should not be regulated as this is normal type trading, it's detectable by humans and there is no problem there and 500ms rule won't be a problem for them.

This rule only won't solve all the problems in US equity markers, other issues must be handled, such as dark pools, sub penny trading, flash orders, marker making rule enforcement and increase in internalization of trades.

There are 4 main problems in US equity markets: sub second HFT, dark pools, sub penny trading and non reliable market makers.

Dark Pools: appeared as a counter measure to predatory hft trading, where large size traders could not reliably make transactions on nasdaq or nyse and basically went under the carpet to do it, away from hft polluted markets. However this poses a big problem – dark pools don't participate in price discovery mechanism which is the main function of financial markets and thus violate the most basic principle of fair open market – namely all market participants must display price and size. The core of problem is again in sub second hft predatory trading, once that problem is solved and quotation system is cleaned from toxic hft trading then expect large size to come back into the market, but also SEC should force dark pools to display prices and sizes for the market to be fair and open. Else there is a 2 tiered market, one with large size that does not participate in price discovery and another thinner market that smaller traders use and that can easily be manipulated with price impact to the benefit of dark pool traders. For instance you can easily buy large size

lot on a dark pool without any market impact whatsoever and then go into the main thin market and easily whack it in your favor with price impact algo, while unloading into the move and thus generate huge profits, arbing between the dark pool and main markets. In this case one market is used for positioning and other for manipulation with no information factoring whatsoever and no price discovery needed. Dark pools are blatant violation of most basic trading rules: not displaying price and size, not participating in price discovery mechanism.

Sub penny trading: should be banned except for penny stocks. Sub penny trading opens another loophole which hft exploit on a regular basis – ability to front run and hijack best bid or best offer without being seen on the quotation system. How can SEC allow this loophole to exist at all? How can you allow trading to take place at sub penny levels but do not allow displaying the sub penny levels? This is exactly what these firms want! They don't want to be seen sub penny trading but they want to trade on those scales and they don't care if you display the trade price after it already happened. As a result regular traders and investors don't see the real market bids and offers. It creates hidden micro market inside best bid and offer that only these hft firms can see. This is unbelievable, again a blatant violation of most basic trading rule: true quote prices must clearly be displayed. SEC should ban sub penny BOTH quoting and trading without any bypass rules or exemptions for anybody.

There are 3 main loopholes that sub second hft firms exploit:

1. flash orders or any type of order preview, latency based or for sale allows them to become first in line to use the order flow, either to front run it, tail gate it or take it
2. lack of minimum quote exposure rule allows them to “zoom in” on very fine **sub second time scales** and thrive there
3. weak sub penny rule allows them to “zoom in” on very fine **sub penny price scales** and thrive there

**Cumulatively it creates microscopic price-time scales ( sub pennies vs milliseconds ) which human traders/investors can not effectively operate on!!! Thus trading must be brought back closer to the human scales: pennies and seconds.**

Market makers: routinely not honoring their quotes, it gets ridiculous to point of becoming useless of trying to execute against them. The question is then where the liquidity that they are supposed to provide ? It seems they only provide it in rare cases when they need it for themselves. In fast markets they abandon their duties and pull away from the inside market to a safe distance. SEC should design the rules in such a way so that market makers can not pull out a quote if someone is executing against their quote. They must honor their quotes as everybody else is required to do, else don't put out an empty quote. Another solution is to abandon the concept of market making altogether because simply is not working and the concept is obsolete, in this case remove any market maker benefits from them and let them trade as everybody else does. Electronic Future exchanges don't have any market makers and market functions fine without them. Basically in present form Market Maker rules work only for the benefit to them and not the markets, it's a loophole exemption they exploit.

Other issues that are less important:

Short bans:

creates biases in favor of traders with long positions, whereas shorting is healthy for the market. If a stock is attacked by the shorts then it's the problem with the stock, it's weak and is being shorted and if information is being manipulated about the stock then stock will return to true value after information is corrected, which will result in a short squeeze. Shorting is not without a risk. Basically looking at a selective short ban as a defense mechanism against attacks by the shorts is not a valid argument. Shorts attack on information and by factoring information and if that information is incorrect then they will be squeezed hard in the short squeeze, thus the risk is large when shorting. Instead of short bans company officials should have contingency plans ready to repudiate the incorrect information on air and those plans should be ready for deployment in case they think their stock is unfairly attacked by the shorts, and if a CEO is not on air and not defending the stock in the open when stock is being heavily shorted then it's a clear sign of weakness and the stock deserves being shorted as true value needs to be found. Else if you allow selective short bans you risk the case where any CEO can call the SEC to get the short ban enabled before information can be factored by the market forces, this does not sound right and goes counter the notion that the Market should determine the true value of stock by factoring information and not the CEO's call to the government SEC.

Circuit breakers:

In the recent flash crash, there was a gap in the circuit breaker rules, that gap must be closed with rules that apply to times from market open to market close as market crash can happen within 10 minutes. During the recent flash crash the market covered 500 Dow points movement within 5 minutes, meaning market can crash 10 minutes before the close. Circuit breaker rules must not change frequently and must be clearly posted on the SEC web page with update notes.

Internalization: increasing number of trades are happening internally away from the problem ridden main markets. Fixing hft problems coupled with banning sub penny trading and normalizing dark pool, should help bringing trading back into the main markets.

In short, to fix the problem in US equity markets:

1. Enable minimum quote exposure rule - 300ms to 1 second, one rule for everybody across all equity markets, without exemptions or loopholes
2. Get rid sub penny trading/quoting altogether, except for penny stocks
3. Normalize dark pools, make them come out from under the carpet and force them to participate in price discovery.
4. Do not enable any mechanism for selective or aggregate short bans, up tick rules. Let the market do its job at factoring information. Make CEOs responsible for defense against shorting.
5. Close time gaps in circuit breaker rules and simplify them.