The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response

Eric Budish, Peter Cramton and John Shim

Inaugural CAFIN Workshop

25 April 2014
The HFT Arms Race: Example

- In 2010, Spread Networks invests $300mm to dig a high-speed fiber optic cable from NYC to Chicago.
- Shaves round-trip data transmission time ... from 16ms to 13ms.
- Industry observers: 3ms is an “eternity”.
- Joke at the time: next innovation will be to dig a tunnel, “avoiding the planet’s pesky curvature”.
- Joke isn’t that funny ... Spread’s cable is already obsolete!
- Not tunnels, but microwaves (first 10ms, then 9ms, now 8.5ms).
- Analogous races occurring throughout the financial system, sometimes measured as finely as microseconds or nanoseconds
  - Last month alone: Lasers, BusinessWire
The HFT Arms Race: Market Design Perspective

- We examine the HFT arms race from the perspective of market design.
  - We assume that HFT’s are optimizing with respect to market rules as they’re presently given.
  - But, ask whether these are the right rules.
    - Avoids much of the “is HFT good or evil?” that seems to dominate the discussion of HFT.
    - Instead, ask at a deeper level what is it about market design that incentivizes arms race behavior, and is this design optimal.
  - Central point: HFT arms race is a *symptom* of a basic flaw in modern financial market design: continuous-time trading.
  - Proposal: make time discrete.
    - Replace continuous-time limit order books with *discrete-time frequent batch auctions*: uniform-price sealed-bid double auctions conducted at frequent but discrete time intervals, e.g., every 1 second or 100ms.
    - Minimal departure from CLOB that eliminates the problems caused by continuous-time trading.
Frequent Batch Auctions

A simple idea: make time discrete.

1. Continuous limit-order books don’t actually “work” in continuous time: market correlations completely break down; frequent technical arbitrage opportunities

2. Technical arbitrage opportunities \(\rightarrow\) arms race.

3. Theory model: critique of the CLOB market design
   - Shows that the arms race is a never-ending, equilibrium, feature of the market design (helps explain empirical facts)
   - Identifies the costs of the arms race
     - Harms liquidity (spreads, depth)
     - Socially wasteful

4. Frequent Batch Auctions as a market design response
   - Benefits: eliminates arms race, enhances liquidity, computational advantages
   - Cost: investors must wait a small amount of time to trade
Frequent Batch Auctions

A simple idea: make time discrete.

1. Continuous limit-order books don’t actually “work” in continuous time: market correlations completely break down; frequent technical arbitrage opportunities
2. Technical arbitrage opportunities \(\rightarrow\) arms race.
3. Theory model: critique of the CLOB market design
   - Shows that the arms race is a never-ending, equilibrium, feature of the market design (helps explain empirical facts)
   - Identifies the costs of the arms race
     - Harms liquidity (spreads, depth)
     - Socially wasteful
4. Frequent Batch Auctions as a market design response
   - Benefits: eliminates arms race, enhances liquidity, computational advantages
   - Cost: investors must wait a small amount of time to trade
Market Correlations Break Down at High Frequency

ES vs. SPY: 1 Day

ES Midpoint
SPY Midpoint
Market Correlations Break Down at High Frequency

ES vs. SPY: 1 hour
Market Correlations Break Down at High Frequency

ES vs. SPY: 1 minute
Market Correlations Break Down at High Frequency

ES vs. SPY: 250 milliseconds
Market Correlations Break Down at High Frequency
ES vs. SPY: Correlations by Time Interval in 2011

(a) 30 seconds
(b) 100 milliseconds
## Correlation Breakdown = Technical Arb Opportunities

Summary Stats, ES-SPY Arb, 2005-2011

<table>
<thead>
<tr>
<th></th>
<th>Percentile</th>
<th>Mean</th>
<th>1</th>
<th>5</th>
<th>25</th>
<th>50</th>
<th>75</th>
<th>95</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong># of Arbs/Day</strong></td>
<td></td>
<td>801</td>
<td>118</td>
<td>173</td>
<td>285</td>
<td>439</td>
<td>876</td>
<td>2498</td>
<td>5353</td>
</tr>
<tr>
<td><strong>Profits/Arb (Idx Pts)</strong></td>
<td></td>
<td>0.09</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.08</td>
<td>0.11</td>
<td>0.15</td>
<td>0.22</td>
</tr>
<tr>
<td><strong>Profits/Day ($)</strong></td>
<td></td>
<td>306k</td>
<td>27k</td>
<td>39k</td>
<td>75k</td>
<td>128k</td>
<td>218k</td>
<td>756k</td>
<td>2333k</td>
</tr>
</tbody>
</table>
Frequent Batch Auctions

A simple idea: make time discrete.

1. Continuous limit-order books don’t actually “work” in continuous time: market correlations completely break down; frequent technical arbitrage opportunities

2. **Technical arbitrage opportunities** $\rightarrow$ **arms race**.

3. Theory model: critique of the CLOB market design
   - Shows that the arms race is a never-ending, equilibrium, feature of the market design (helps explain empirical facts)
   - Identifies the costs of the arms race
     - Harms liquidity (spreads, depth)
     - Socially wasteful

4. Frequent Batch Auctions as a market design response
   - Benefits: eliminates arms race, enhances liquidity, computational advantages
   - Cost: investors must wait a small amount of time to trade
Arb Durations over Time: 2005-2011

(c) Median over time

(d) Distribution by year
Arb Per-Unit Profits over Time: 2005-2011

(e) Median over time

(f) Distribution by year
Arb Frequency over Time: 2005-2011

(g) Median over time

(h) Frequency vs. Volatility
Arms Race is a “Constant” of the Market Design

- Results suggest that the arms race is a mechanical “constant” of the continuous limit order book.
  - Rather than a profit opportunity that is competed away over time

- Correlation Breakdown
  - Competition does increase the speed with which information is incorporated from one security price into another security price
  - Competition does not eliminate correlation breakdown

- Technical Arbitrage
  - Competition does increase the speed requirements for capturing arbs (“raises the bar”)
  - Competition does not reduce the size or frequency of arb opportunities

- These facts both inform and are explained by our model
Total Size of the Arms Race Prize

- Estimate annual value of ES-SPY arbitrage is $75mm (we suspect underestimate, details in paper)
- And ES-SPY is just the tip of the iceberg in the race for speed:

1. Hundreds of trades very similar to ES-SPY: highly correlated, highly liquid
2. Fragmented equity markets: can arbitrage SPY on NYSE against SPY on NASDAQ! Even simpler than ES-SPY.
3. Correlations that are high but far from one can also be exploited in a statistical sense. Example: GS-MS
4. Race to top of book (artifact of minimum tick increment and maker-taker)
5. Race to respond to public news (eg Business Wire, Fed)

We don’t attempt to put a precise estimate on the total prize at stake in the arms race, but common sense extrapolation from our ES-SPY estimates suggest that the sums are substantial.
Frequent Batch Auctions

A simple idea: make time discrete.

1. Continuous limit-order books don’t actually “work” in continuous time: market correlations completely break down; frequent technical arbitrage opportunities
2. Technical arbitrage opportunities → arms race.
3. Theory model: critique of the CLOB market design
   - Shows that the arms race is a never-ending, equilibrium, feature of the market design (helps explain empirical facts)
   - Identifies the costs of the arms race
     - Harms liquidity (spreads, depth)
     - Socially wasteful
4. Frequent Batch Auctions as a market design response
   - Benefits: eliminates arms race, enhances liquidity, computational advantages
   - Cost: investors must wait a small amount of time to trade
Key idea: in a CLOB, any time there is public information, there is a race. This race ultimately harms liquidity provision.

Why? Consider the race from a liquidity provider’s perspective

- Suppose there is a publicly observable news event that causes his quotes to become “stale”
  - E.g., a change in the price of a highly correlated security (ES/SPY), Fed announcement
- 1 of him, trying to adjust his stale quotes
- Many others, trying to “snipe” his stale quotes
- In a continuous limit order book, messages are processed one-at-a-time in serial ...
- so the 1 usually loses the race against the Many ...
- Even if he, too, is at the cutting edge of speed

Incentive to be fast

- Snipers: win race to pick off stale quotes
- Liquidity providers: get out of the way of the snipers!
Model: Key Idea

- This technical cost of providing liquidity – liquidity-providing HFTs getting “sniped” by other HFTs in the race to respond to symmetrically observable public news – is incremental to the usual fundamental costs of providing liquidity
  - Asymmetric information, inventory costs, search costs
- In a competitive market, picking off costs get passed on to investors
  - Thinner markets, wider bid-ask spreads
- Ultimately, in equilibrium of our model, all of the $ spent in the arms race come out of the pockets of investors
  - Arms-race prize = expenditures on speed = cost to investors
  - Remember: arms-race profits have to come from somewhere
Model: Additional Remarks

The Arms-Race is a “Constant”

- Comparative static: the negative effects of the arms race do not depend on either
  - the cost of speed (if speed is cheap, there will be more entry)
  - the magnitude of speed improvements (seconds, milliseconds, microseconds, nanoseconds, ...)

- The problem we identify is an equilibrium feature of continuous limit order books
  - not competed away as HFTs get faster and faster
  - ties in nicely with empirical results

- Takeaway: the race for speed will never end as long as we have continuous-time trading
Role of HFTs

In our model HFTs endogenously perform two functions:

- Useful: liquidity provision / price discovery
- Rent-seeking: picking off stale quotes

HFTs are indifferent between these two roles in equilibrium of our model.

The rent-seeking seems like zero-sum activity among HFTs:

- but we show that it ultimately harms real investors

Frequent batching preserves the useful function but eliminates the rent seeking function (or at least reduces)
What’s the Market Failure?

Chicago question: isn’t the arms race just healthy competition? what’s the market failure?
What’s the Market Failure?

Our model yields two responses

1. Model shows that the arms race can be interpreted as a prisoners’ dilemma
   ▶ If all HFTs could commit not to invest in speed, they’d all be better off
   ▶ But, each individual HFT has incentive to deviate and invest in speed

2. Model shows that a violation of the efficient market hypothesis is built in to the market design
   ▶ Violations of the the weak-form EMH are intrinsic to the CLOB
   ▶ You can make money from purely technical information (and HFTs do!)
   ▶ Core issue: continuous markets process messages in serial (i.e., one-at-a-time)
   ▶ Even for public / technical info (e.g., a jump in ES): somebody is always first to react
A simple idea: make time discrete.

1. Continuous limit-order books don’t actually “work” in continuous time: market correlations completely break down; frequent technical arbitrage opportunities

2. Technical arbitrage opportunities → arms race. Arms race does reduce the duration of the market failures; arms race does not actually fix the market failures

3. Theory model: critique of the CLOB market design
   ▶ Shows that the arms race is a never-ending, equilibrium, feature of the market design (helps explain empirical facts)
   ▶ Identifies the costs of the arms race
     ▶ Harms liquidity (spreads, depth)
     ▶ Socially wasteful

4. Frequent Batch Auctions as a market design response
   ▶ Benefits: eliminates arms race, enhances liquidity, computational advantages
   ▶ Cost: investors must wait a small amount of time to trade
Frequent Batch Auctions: Overview

- High level: analogous to a CLOB, except time is discrete
- Discrete time then necessitates batch processing, using an auction
Frequent Batch Auctions: Details

Order Submission

- The trading day is divided into equal-length discrete “batch intervals”, each of length $\tau > 0$ (e.g. 1 second)
- During the batch interval, traders submit bids and asks as price-quantity pairs
  - Can be freely modified, withdrawn at any time
  - If an order is not executed in the batch at time $t$, it automatically carries over for $t + 1$, $t + 2$, etc., until it is either executed or withdrawn
  - Analogous to standard limit orders
Frequent Batch Auctions: Details

Auction

- At the conclusion of each batch interval, the exchange “batches” all of the outstanding orders, and computes market-level supply and demand curves
- Case 1: supply and demand don’t cross
  - No trade
  - All orders remain outstanding for the next batch auction
  - We expect this case to be quite common
  - Analogous to liquidity provision in a CLOB
- Case 2: supply and demand cross (at price $p^*$)
  - All bids strictly greater than $p^*$ and all asks strictly lower than $p^*$ transact their full quantity at $p^*$ (“uniform price”)
  - Bids and asks of exactly $p^*$ may get rationed
  - Suggested rationing rule: pro-rata with time priority across batch intervals but not within batch intervals
  - Historical aside: uniform-price auction originally proposed by Milton Friedman in the 1960s, for Treasury auctions.
Frequent Batch Auctions: Illustrated

(a) Case 1: No Trade

(b) Case 2: Trade
Frequent Batch Auctions: Details

Reporting

- After the auction is computed, the following information is announced publicly
  - The market-clearing price (or the outcome “no trade”)
  - The quantity cleared (possibly zero)
  - The supply and demand curves
  - Optional: individual bids comprising the supply and demand curves

- Key point: orders for the time $t$ batch auction are not visible during the time $t$ batch interval. Instead, announced after the time $t$ auction is conducted. This is to prevent gaming.

- Analogous to current practice under the CLOB
  - (i) submit order; (ii) exchange processes order; (iii) public announcement
  - For auction at time $t$, traders see bids and asks at time $t - 1, t - 2, t - 3, \ldots$ (see market as it was a latency ago)
Frequent Batch Auctions: Illustrated
Why and How Batching Eliminates the Arms Race

There are two reasons why batching eliminates the arms race:

1. Batching reduces the value of a tiny speed advantage
   - If the batch interval is 1 second, a 1 millisecond speed advantage is only \( \frac{1}{1000} \)th as useful.

2. Batching transforms competition on speed into competition on price
   - Ex: the Fed announces policy change at 2:00:00.000pm ...
     - Continuous market: competition manifests in a race to react. *Someone is always first.*
     - Batched market: competition simply drives the price to its new correct level for 2:00:01.000. Lots of orders reach the exchange by the end of the batch interval.
Computational Benefits of Frequent Batching

- **Overall**
  - Continuous-time markets implicitly assume that computers and communications technology are infinitely fast.
  - Discrete time respects the limits of computers and communications. Computers are fast but not infinitely so.

- **Algorithmic traders**
  - Continuous: Always uncertain about current state; temptation to trade off robustness for speed
  - Discrete: Everyone knows state at time \( t \) before decision at time \( t + 1 \)

- **Exchanges**
  - Continuous: Computational task is mathematically impossible; latencies and backlog unavoidable
  - Discrete: Computation is easy

- **Regulator**
  - Continuous: Audit trail is difficult to parse; who knew what when? in what order did events occur across markets?
  - Discrete: Simple audit trail; state at \( t, t + 1, \ldots \)
Summary: Costs and Benefits of Frequent Batching

- **Benefits**
  - Enhanced liquidity
    - Narrower spreads
    - Increased depth
  - Eliminate socially wasteful arms race
  - Computational / market stability benefits of discrete time

- **Costs**
  - Investors must wait until the end of the batch interval to transact
  - Transition costs
  - Unintended consequences (but: remember that the continuous market has itself had numerous unintended consequences which batching addresses)
Summary

▶ The root problem isn’t “evil HFTs”
▶ The root problem is the continuous-time, serial-process market design
▶ Boring ...
▶ But simpler to solve
▶ Solution: make time discrete, and batch process

1. Stops the arms race
2. Transforms the nature of competition
   ◀ Competition on speed → Competition on price
   ◀ Incentive to be faster → Incentive to be smarter
   ◀ Preserve the useful function of HFT; Reduce rent-seeking

3. Computationally realistic