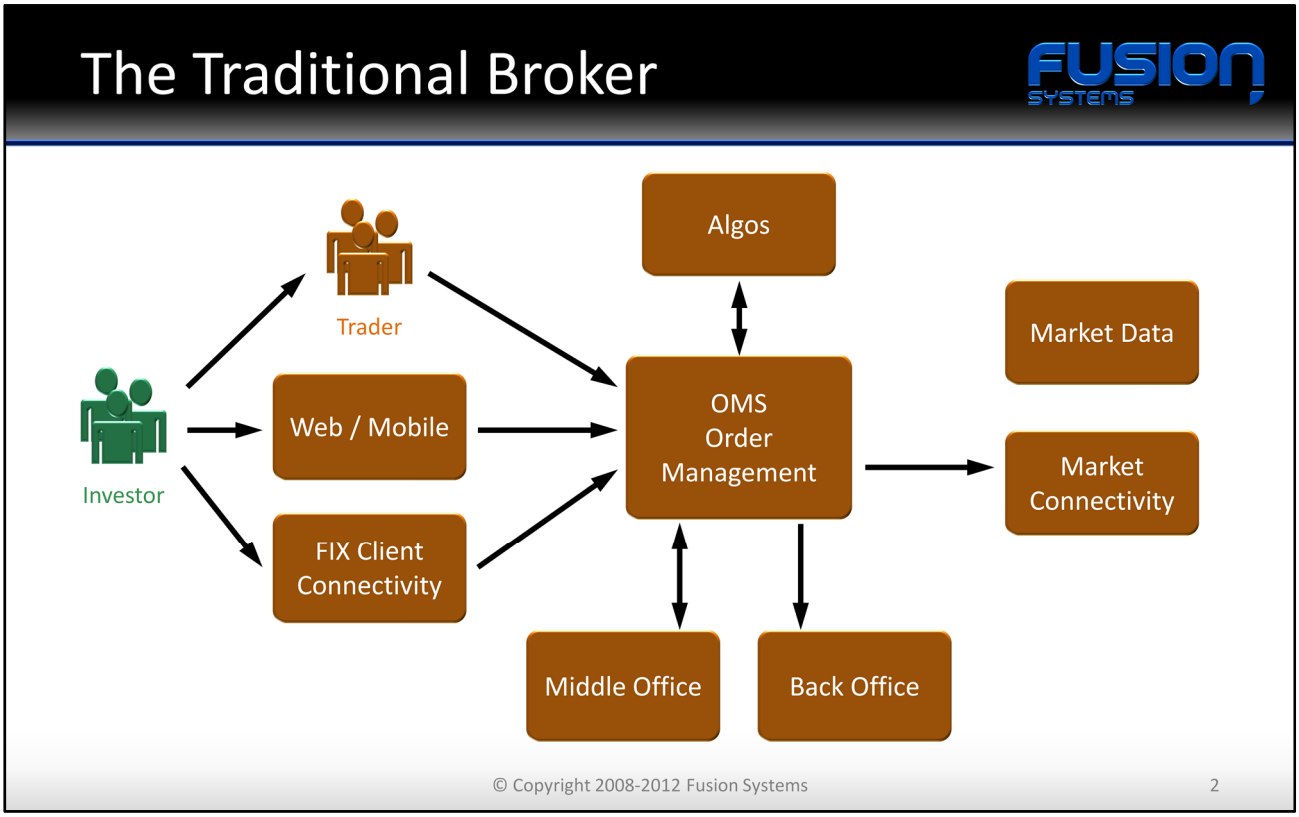
Low Latency Risk Management

Implications of Changing Markets

With this presentation we wanted to share our perspective on the requirements that are driving new trading technology, particularly in APAC, and particularly with reference to the diversity and complexity of market mechanisms in the region.

While latency is a key challenge, risk management is equally if not more challenging, and these in combination with several other elements are key to being able to deploy a premium electronic trading service.



This is a recap of technology on the sell side. These are the major front office technology systems found in traditional brokers.

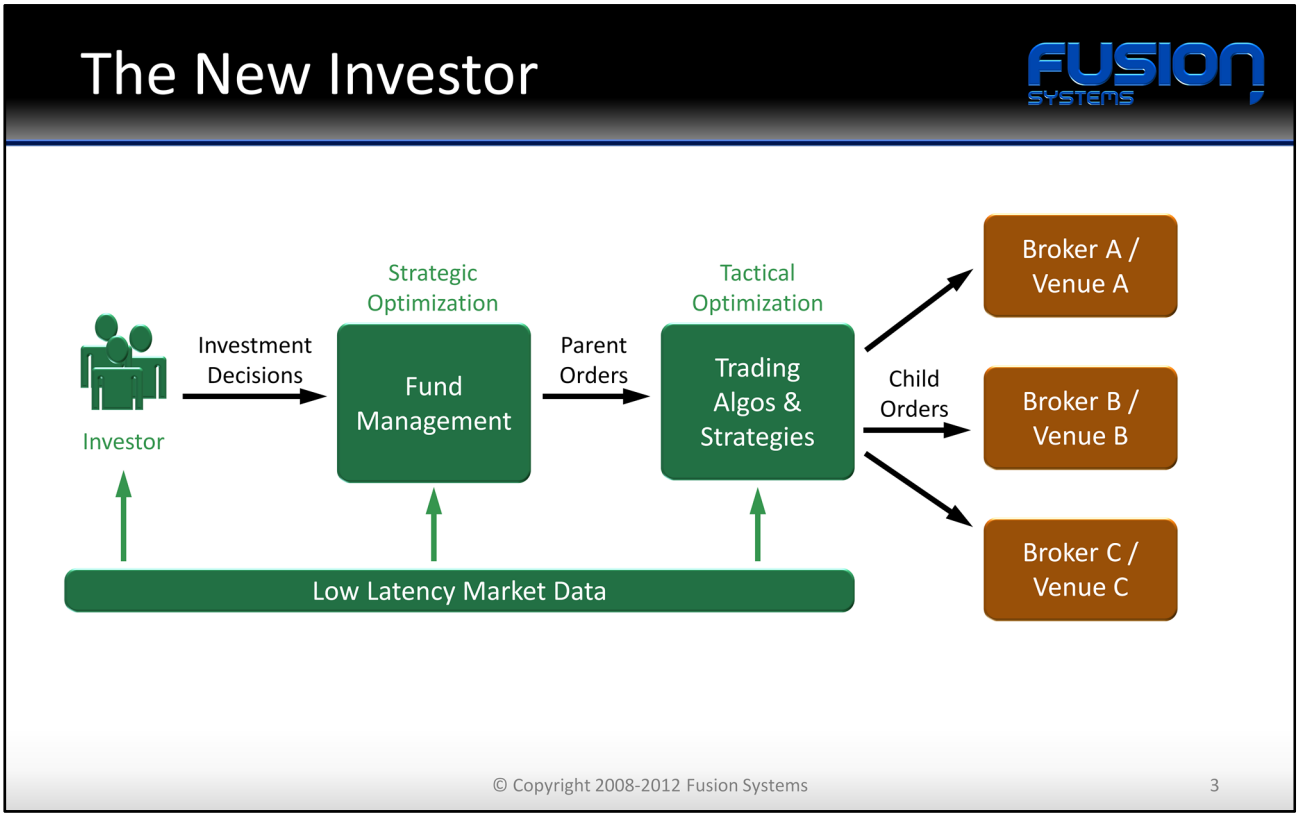
Orders from Investors are communicated to brokers via a variety of channels. Worked orders are handled by traders, and online trading is handled by Web and mobile application servers. Electronic trading is conducted over FIX networks, arriving at the broker's order management system via client connectivity, typically a cluster of FIX engines.

The OMS will then perform pre-trade risk management, often by passing the order to a separate middle office risk system. If the order is one of the standard algos (VWAP, inline, smart order routing, etc.), it will go to an Algo system, possibly provided by a 3rd party or another broker.

Finally marketable orders will flow from the OMS to the market, via market connectivity. This is typically a cluster of market gateways or line handlers that actually link to the market.

From the time the order is initiated and sent by the investor, to the time it arrives at the market, it will have passed through at least three and as many as five separate systems. Elapsed time in transit may be as much as tens or even hundreds of milliseconds.

With the speed at which today's markets change, the order may well be mispriced ,or its' economic basis invalidated, while it is in transit. The longer the transit time, the higher the latency, the more likely it is that trading profits will suffer due to stale and mispriced orders being filled, aka slippage.



A (not so) new generation of investors have completely changed the way that they make investment and trading decisions. The basic approach was pioneered by Jesse Livermore in the 20s, but once joined with electronic trading the trend really took off in the US markets 25 years ago with the “Market Wizards”. It shows no signs of slowing down.

These investors orchestrate a diverse range of automated trading strategies in order to achieve their investment goals. Their investment decisions revolve around optimizing the allocation of funds to different strategies over time.

They devote sustained systematic efforts to designing, implementing, monitoring and tuning these proprietary automated trading strategies.

At any time there is a portfolio of available strategies, some of which will be active. Active strategies are started and stopped, given more or less budget, and have their operating parameters amended through the trading day in response to market conditions and their ongoing investment returns. These commands – start, stop, amend size, amend parameters – are parent orders.

So the output of the strategic optimization process is a series of parent orders that control the running of strategies. The output of the tactical optimization process is a series of child orders that are sent to brokers and venues. These include new orders, amends and cancels.

All of these processes are critically dependent on accurate and up-to-date information about the state of the markets, i.e. they depend on low-latency market data.

The New Investor's Algos

- Market Making
  - Re-price quotes promptly in response to market updates
- Arbitrage
  - Index arbitrage (leaders vs laggers)
  - Basis trading (derivative vs underlying)
- Liquidity Discovery
  - Probing the market with small-sized orders to reveal hidden liquidity
- Order Anticipation
  - Matching and responding to patterns of trading
- Momentum Ignition
  - Precipitating trends by reinforcing developing patterns of trading
- All need rapid order execution in response to market data updates
  - Slippage and stale orders reduce trading performance and profits
- **"Tick to Order Book" latency is key**

VWAP, Inline, Arrival price, etc. are conspicuous by their absence. These are not algos for buy and hold investors.

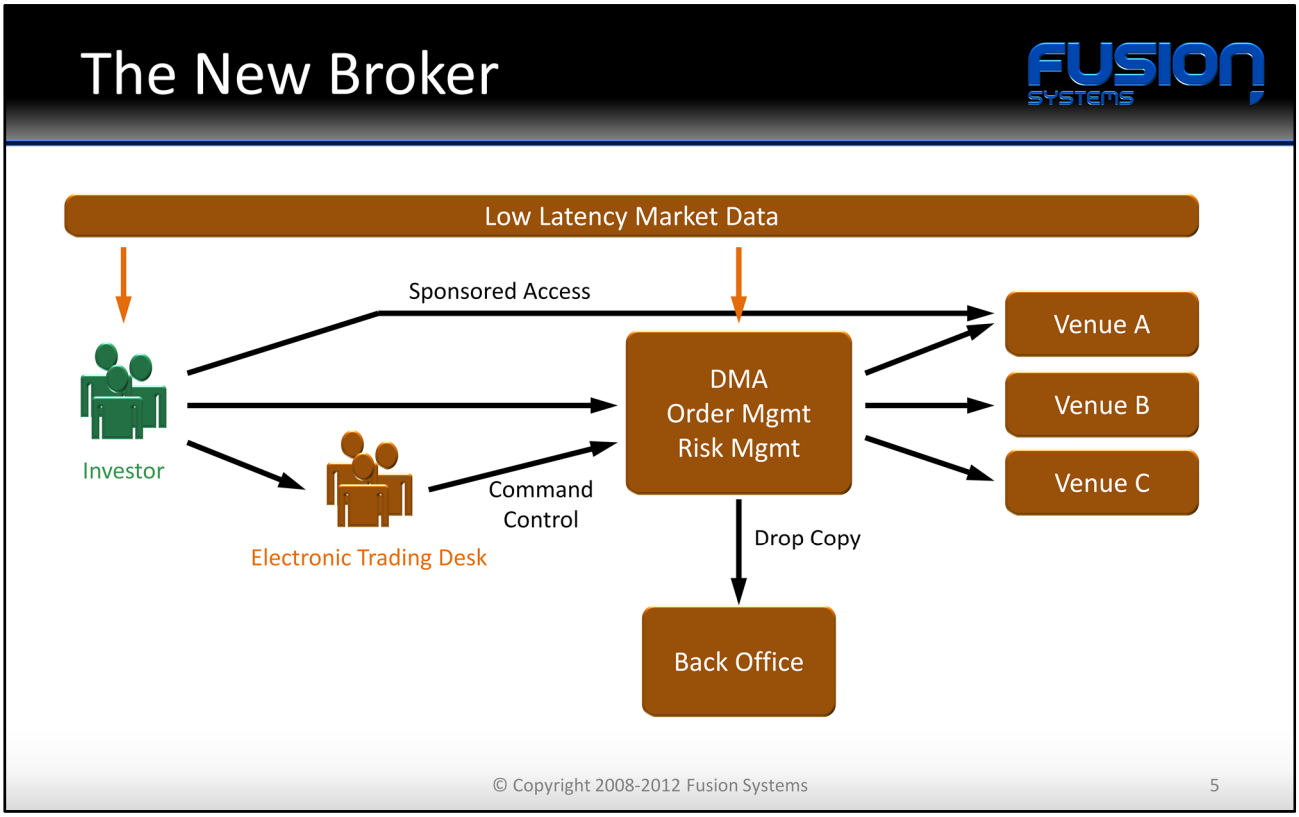
Market making is one of the most common strategies and requires continuous re-pricing of orders in response to market updates, or ticks. Market making will typically benefit from consolidating multiple bids and asks into a single "quote" message, if the market supports that.

Various kinds of arbitrage are popular, particularly lead/lag index arbitrage and basis trading. All such strategies require prompt order execution in response to price moves. For example single stock basis trading will submit orders for the stock in response to price updates for a derivative such as the corresponding single stock future.

Liquidity discovery is the process of posting small quotes that, when filled, reveal reserve liquidity that is not quoted and is otherwise concealed from the lit market. For example, there may be stopped orders or iceberg orders that are held by brokers and will automatically release. By pinging or probing, traders can identify opportunities to quickly go long or short while deferring or reducing the accompanying market impact. Again, orders must be rapidly submitted and re-priced in response to market updates.

Order anticipation and momentum ignition also depend on rapidly responding to market updates – in this case pattern matching of trading activity reported via market data feeds.

For all these types of strategies, "Tick to Order Book" latency is key to trading performance.



To respond to the needs of the new investor, brokers need new services and technology.

Prime brokerage and electronic trading services have become technology-enabled services that hinge upon providing investors with fast reliable access to markets.

In order to ensure the compliance of investor's orders with broker and market restrictions, and prevent market disruption, investor order flows must be validated and risk checked before they are forwarded to markets.

The pre-trade risk checking requirement, previously a middle office function, is now combined with client connectivity, order management and market connectivity into a single DMA platform that provides a unified solution to the broker and the investor. An essential feature of such platforms is the drop copy for downstream processing by back office.

In addition, the electronic trading desk will need the ability to intervene in investors' order flows. For example investors might experience network failure and request the broker to mass cancel open orders on their behalf, or to resend the day's execution reports to the investor to recover from an outage.

For investors seeking the absolute fastest, most direct market access, brokers can provide sponsored access to dedicated market links that are provisioned just for that investor, but specialized network hardware for pre-trade risk checking will generally still be required, since so-called naked access is no longer permitted in many markets and is in any case prohibited by most brokers' internal compliance rules.

The New Broker – Differentiators

- Trading Costs – Synthetic Swap Settlement
- Financing of Trading – Leverage & Margin Collateral Sharing
- Long-Short Facilitation – Inventory Access
- Dark Liquidity Access
- System Latency & Capacity
  - Ultra Low Latency Market Data
  - Ultra Low Latency Order Execution
  - Fast, Transparent Pre-Trade Risk Checks
- Tick by Tick Historical Data for Back-Testing

Brokers can attract these new investors with new services and value propositions.

Business differentiators include reducing trading costs, for example by offering synthetic swap settlement; financing trading; facilitating long and short selling; and providing access to dark liquidity.

While primarily business value propositions, none of these options can be offered to investors without specialized support built-in to the DMA trading platform.

Fast market connectivity is not enough!

Naturally, the trading platform provided to the investor must satisfy the basic prerequisites of ultra low latency market data, ultra low latency order execution and pre-trade risk checking.

To complete the service offering, the broker should facilitate back testing of automated trading strategies by offering Level 2 tick by tick historical data for the supported venues.

Electronic Trading Services

- Connectivity Management
  - Client Links
  - Market Links
- Order/Execution Flow Management
  - Interventions – Trader Support & Facilitation
  - Active Risk Management – Risk Alerts & Limit Changes
  - Handle Trading Disruptions – Disable Flow and Mass Cancel
- Market Data Management
- Client Onboarding, Testing

Internally the broker must organize itself to provide a premium level of electronic trading services.

Hundreds of thousands of orders flow at very high rates from multiple investors to multiple venues in this environment. In order to be able to manage such a high performance electronic trading environment, highly productive teams are essential.

Those teams need real-time visibility, command and control of all the elements of the electronic trading infrastructure. They need to be able to manage risk, respond to alerts, troubleshoot problems, support investors and intervene in trading as needed.

The only way to achieve the required levels of productivity is by providing the trading desks and trading technology infrastructure teams with sophisticated apps that are integrated with the DMA platform.

Low Latency Risk Management

- Order Pre-Trade Validations
  - Valid Venue, Symbol
  - Valid Price - Round Tick
  - Valid Qty - Round Lot
  - Valid Side, Order Type, Time In Force
  - Valid Order Restrictions, Capacity
- Order Pre-Trade Risk Filters
  - Size - Max Lots, Max Shares, Max %ADV, Max %Outstanding
  - Size / Price - Max Notional Value
  - Price - Price Range from Base/Open/Last
  - Restricted Stock List
  - Stock Short Sell Locate - Ensure Sufficient Stock Loan
  - Stock Long Sell - Ensure Sufficient Stock Inventory
  - Derivatives - Trading Power - Ensure Sufficient Margin Collateral

A big part of the broker's responsibility is managing pre-trade risk checks to maintain compliance, and prevent market disruption and excessive exposure by investors.

Key requirements are accurately tracking inventories for long and short selling of stocks, and margin collateral for derivatives trading.

Many markets have strict compliance requirements for these processes and brokers cannot permit orders to go to market without having passed these detailed very specific checks.

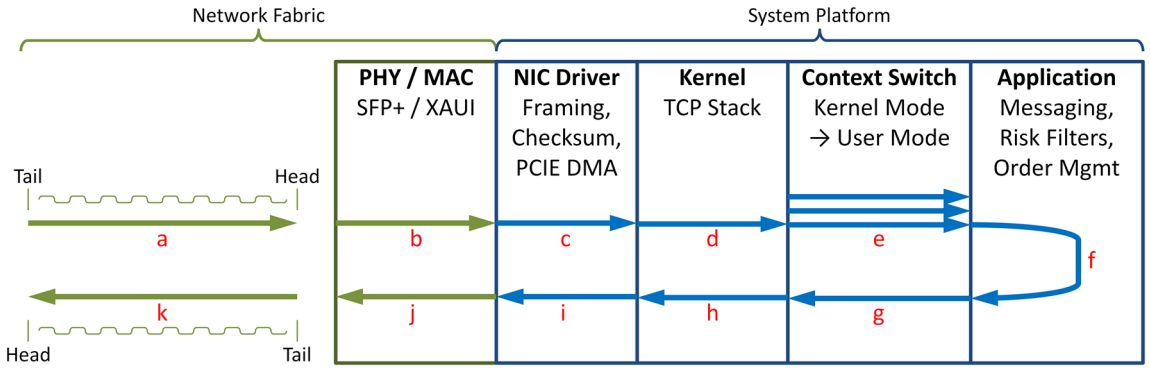
Typical Order Execution Latencies

- Legacy Infrastructures: 2 – 100ms

- FIX DMA: 50 – 200µs
- Binary DMA: 25 – 50µs
- Sponsored Access (FPGA): 1.5 – 8µs

Given the client connectivity, risk checks and market connectivity requirements, here are some typical latencies for the various kinds of systems that brokers can use.

Elements of Latency – x86 Linux



a, k 1GBE – (#bytes + 40) * .008µs
10GBE – (#bytes + 40) * .0008µs
Typical New Order message is 2-3µs over 1GBE

b, j ~0.7µs

c, i 0.5 – 1µs

d, h 0.5 – 1µs

e 1 – 10µs (3 times – epoll_wait, recv, recv)

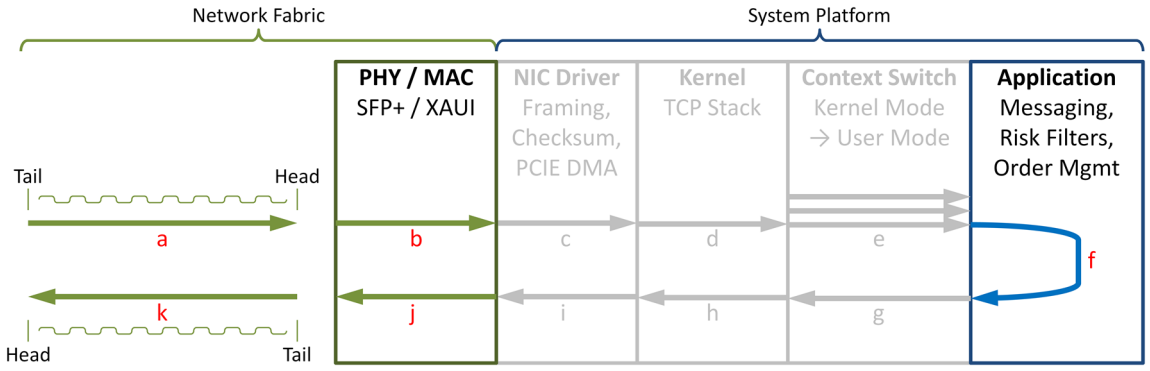
g 1 – 10µs (send)

f 5 – 15µs for binary, 20 – 30µs for FIX

Example Total (10GBE) = **40-60µs**

The latency of regular software applications running on Linux is in large part due to the constraints of the system architecture. Equivalent performance constraints apply to Windows.

Elements of Latency – FPGA



a, k 1GBE – (#bytes + 40) * .008µs
10GBE – (#bytes + 40) * .0008µs
Typical New Order message is 2-3µs over 1GBE
b, j ~0.7µs

f 1 – 1.5µs

Example Total (10GBE) = **1.25µs**

Here we can see the enormous reduction in latency that can be provided by a dedicated FPGA platform by eliminating the overheads introduced by PCIE, the driver and the operating system.

A Broker in 9 Rack Units as a Service

1U Ultra Low Latency Switch

2U Risk-Filtered DMA Gateway with FPGA Card (Primary)

2U Risk-Filtered DMA Gateway with FPGA Card (Backup)

2U ULL Market Data Server

2U Buy-Side Managed Server Automated Trading System

This shows how a market-leading electronic trading service can be provided as a service with as few as 9 rack units, including the investor's co-located server that is hosting their automated trading strategies.

For HKEx we would add one rack unit per OG for the cash market and/or one rack unit per NG for the derivatives market.

Fusion HKEx Ecosystem Services

- Ultra Low Latency Risk-Managed Order Execution
  - FIX DMA
  - Binary DMA
  - Sponsored Access (Native FPGA-based)
  - Stock & Derivatives
- Ultra Low Latency Market Data
- Security Reference Data & Historical Market Data
- Algo / Strategy Managed Hosting for Automated Trading

These are the ecosystem services Fusion will provide from our rack space at HKEx. This is everything an exchange participant needs to provide a premium ultra-low-latency electronic trading service, except for booking, clearance and settlement.

Thank You!

FUSION SYSTEMS

Your Strategic Technology Partner

Tokyo
Daiwa Azabudai Bldg 8F
Azabudai 2-3-3
Minato-ku, 106-0041
Tel: +81 3 3505 2880

Hong Kong
603 Tower 2 Lippo Centre
89 Queensway
Tel: +852 2530 1938

Shanghai
14B1 Catic Bldg
212 Jiang Ning Rd, 200041
Tel: +86 21 5289 5858

www.ulldma.com