

*2011 High Performance Computing  
Low Latency, Throughput and  
Data Center Efficiency*



**Thomas Fay, VP – Engineering NASDAQ OMX**

**April 4, 2011**

**NASDAQ OMX<sup>®</sup>**

# NASDAQ OMX – Get to Know Us

- We invented electronic trading 40 years ago and are now the world's largest exchange company.
- Electronic trading is trusted and emulated by every electronic equities exchange in the world, making the world's capital markets move faster, more efficiently and more transparently.



We power  
**1 in 10**  
of the  
world's  
securities  
transactions

NASDAQ OMX  
trading technology is  
used to power  
**more than 70**  
exchanges  
in **50**  
countries

NASDAQ OMX lists  
**3,600**  
global companies  
worth \$5.4T in  
market cap  
representing diverse  
industries and many  
of the world's most  
well-known and  
innovative brands

Our global platform  
can handle  
**more than 1M**  
messages/second  
at sub-100  
**microsecond**  
average speeds

We own and  
operate  
**24** markets  
**3** clearing houses  
**5** central securities  
depositories

More than  
**\$350B**  
is tied to  
our global indexes

# Why is latency important?

## Low Latency Attracts Order Flow

- High Execution Quality
- Deterministic Performance
  - Low Jitter (Regardless Of Throughput)

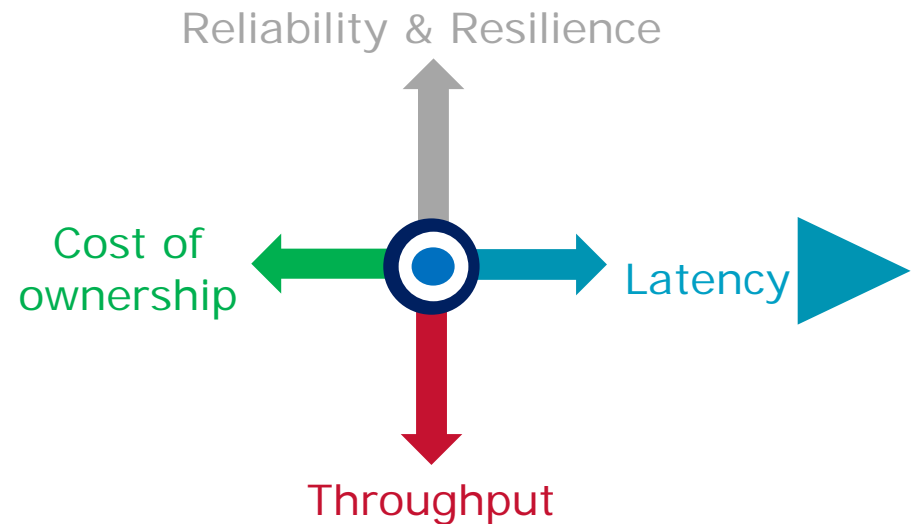


## Competitive Edge

- Exchange Customers Direct Flow Based Upon Best Experience
- Liquidity Begets Liquidity Begets Liquidity!

## Customer Performance Demands Are System Wide

- Fast Acknowledge Times
- High Fidelity/Low Latency Market Data
- Sufficient Throughput To Handle Traffic Bursts
- High Reliability/Up-Time
- Real Time Order Lifecycle Information

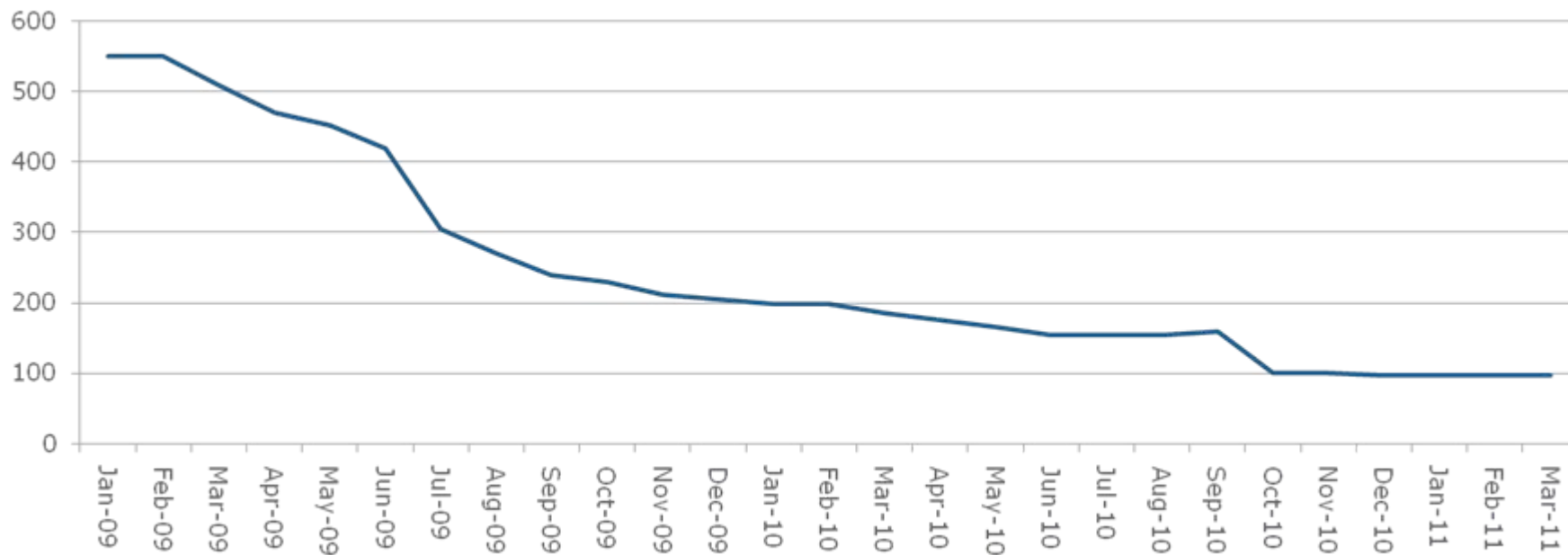


# The Latency Challenge

## NASDAQ MARKET CENTER RECORDS

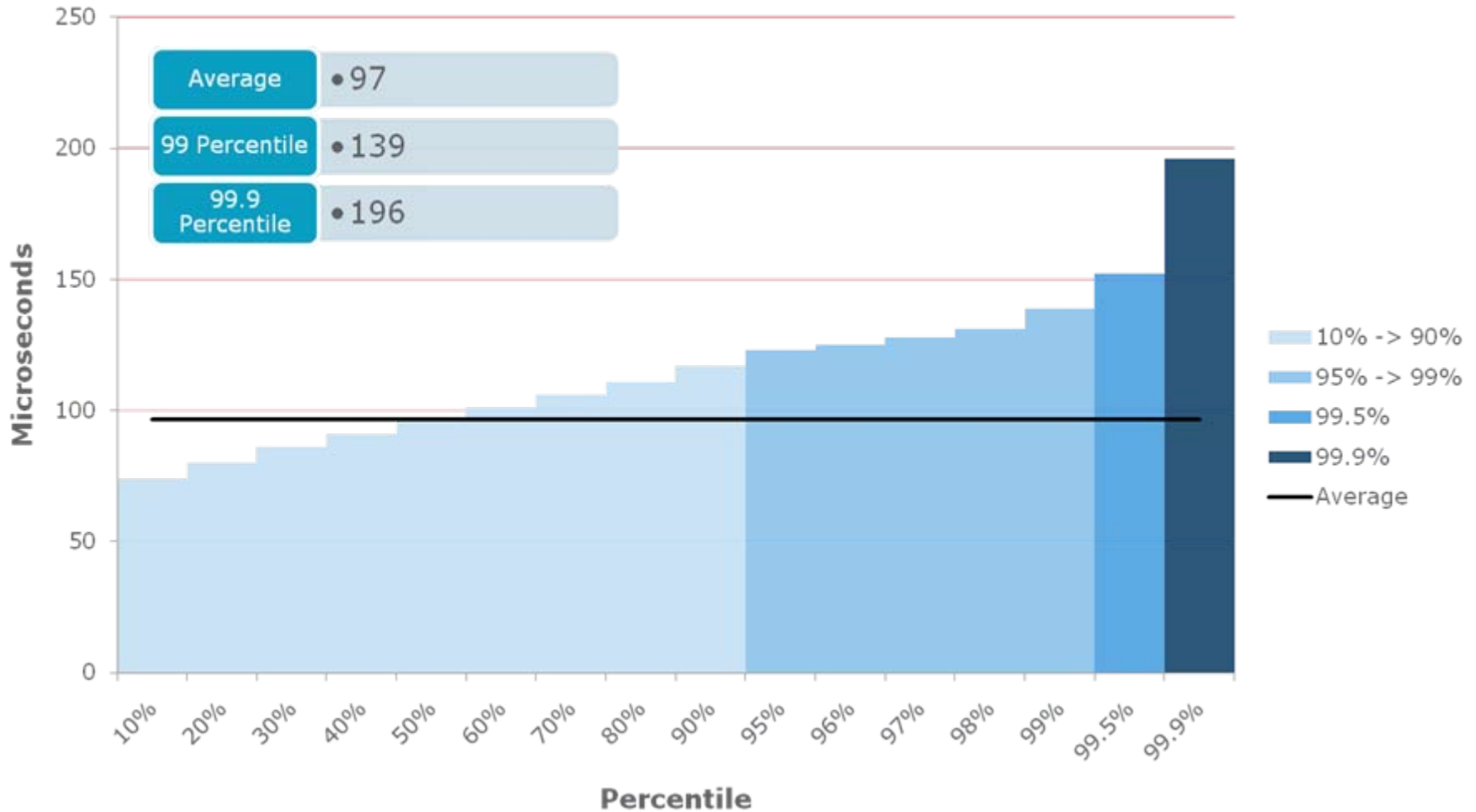
Peak day		Peak second	
Message volume	1,684,103,265	Messages	656,991
Order volume	821,808,375	Orders	226,484
Share volume	12,814,454,760	Executions	117,051

## NASDAQ PERFORMANCE DEVELOPMENT 2009-2011 (Average latency in microseconds)



# 10G Co-Lo Order Latency Summary

## Week of March 21, 2011 (NASDAQ Stock Market 09:30 to 16:30)



# Performance Trends

## Clock Speed "Free-Ride" Is Over

- Concurrency Is Required To Increase Software performance
- Disciplined System Engineering

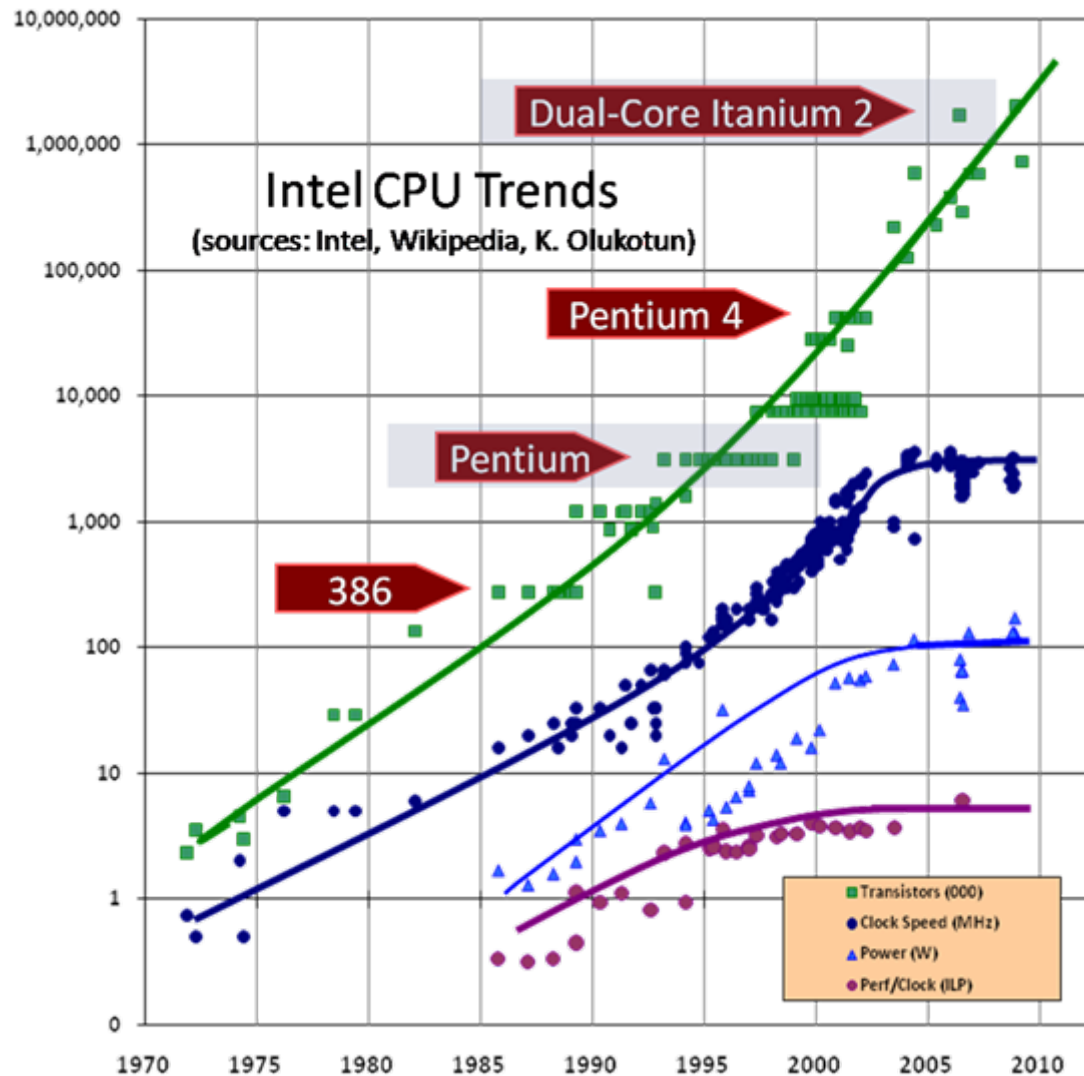
## Multiple Applications Per Server

- Efficient Performance
- Applications Communicate On Server Rather Than Network
- Memory Access Bottlenecks Are Minimized
- Core Pinning And Interrupt Binding

## Continual Decrease In Cost Per Application

- Server Count Reduction From Increase In CPU Core Density
- Lower Power CPUs = Lower Cooling Costs
- Less Servers = Lower Data Center Costs
- Favorable Economics/Attractive TCO

Graph © Herb Sutter, used with permission  
<http://www.gotw.ca/publications/concurrency-ddj.htm>



# The Race to Zero

## Requirement For High Performance And Scalability

- Architectural And Design Considerations
- Agility, Flexibility, Time To Market
- Vigilance – Inter And Intra Server

## Addressing latency in the Network

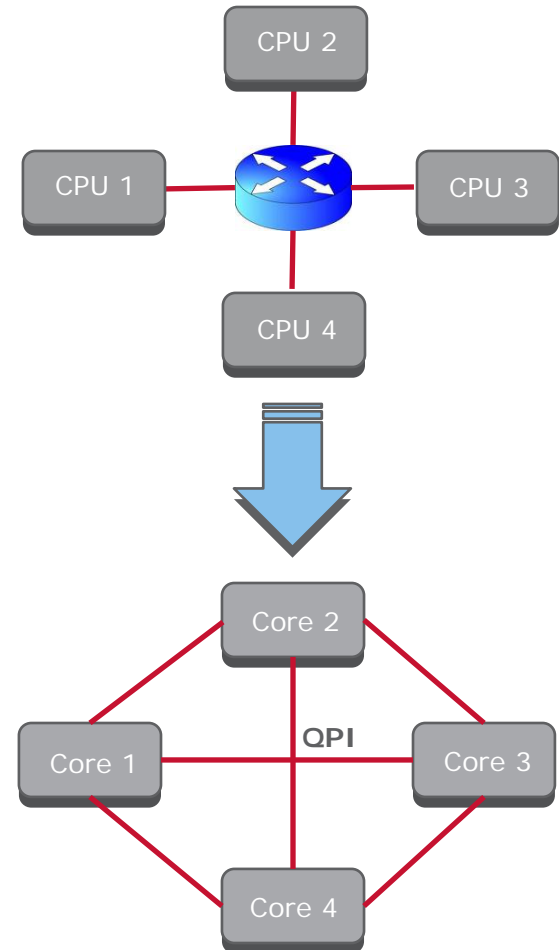
- TCP/IP Bypass
- Switch And Router Optimization
- RDMA And Hardware Acceleration/Appliances
- Highly Deterministic Networking Equipment

## Addressing latency in the Server

- Increasingly Sophisticated Micro-architectures
- Multi-core Technology
- More I/O Logic On CPU Die
- Ultra High Speed Interconnects

## Addressing latency in the Software

- Engineering For Concurrency
- Profiling, Tuning, And Optimization
- RTOS To Increase Determinism
- High Performance Middleware





Thank you.

**NASDAQ OMX<sup>®</sup>**