



UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
VALPARAÍSO – CHILE



HIGH FREQUENCY TRADING & GRAPHICS PROCESSING UNIT

Tesis presentada como requerimiento parcial
para optar al grado académico de

INGENIERO CIVIL

por

Jonathan Andrés Antognini Cavieres

Comisión Evaluadora:

Dr. Luis Salinas Carrasco (Guía, UTFSM)

Dr. Héctor Allende Olivares (UTFSM)

MARZO - 2013

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
VALPARAÍSO – CHILE

TÍTULO DE LA TESIS:

HIGH FREQUENCY TRADING & GRAPHICS PROCESSING UNIT

AUTOR:

JONATHAN ANDRÉS ANTOGNINI CAVIERES

Tesis presentada como requerimiento parcial para optar al grado académico de **Ingeniero Civil** de la Universidad Técnica Federico Santa María.

Profesor Guía:

Dr. Luis Salinas Carrasco

Profesor Correferente

Dr. Héctor Allende Olivares

Marzo - 2013.
Valparaíso, Chile.

A mi familia

Índice general

1. Introducción	1
1.1. Mercados Financieros	2
1.1.1. Mercados bursátiles	3
1.2. Series de tiempo financiera	4
1.2.1. Descripción de serie financiera	5
1.2.2. High frequency trading	6
1.3. High Performance Computing	7
1.3.1. Graphics Processing Unit	7
1.4. Motivación	8
1.5. Organización de la Memoria	9
2. Estado del arte	10
2.1. High Frequency Trading	10
2.2. Wavelet Multiscale	10
2.3. Artificial Neural Network	10
2.4. Graphics Processing Unit	10
3. Descripción del problema	11
3.1. Descripción histórica	11
3.2. Descripción y formalización a usar	11
4. Solución propuesta	12
4.1. Algoritmo y fundamentos teóricos	12

5. Estudio experimental	13
5.1. Selección de data (sector, frecuencia)	13
5.2. Parámetros del algoritmo	13
5.3. Validación	13
5.4. Speed up: multicore, gpu, híbrido	13
6. Conclusiones	14
Bibliography	15

Capítulo 1

Introducción

A través de los años el hombre ha presentado un cambio radical en su nivel de vida; los conocimientos que ha logrado acumular y aplicar han sido para su beneficio, y han cambiado radicalmente su modo de vivir. Existe una notable diferencia entre el hombre de hace unas cuantas décadas y el hombre moderno. Tal diferencia se ha dado por el desarrollo de la ciencia, que está estrechamente relacionada con la innovación tecnológica. Por esta razón se amplía el contenido de cómo ha evolucionado la ciencia y la tecnología en el mundo, su origen remoto, los países que más han aportado en esta área y su respectiva utilización, bien sea para el desarrollo o la destrucción. Como se sabe, la tecnología se está haciendo presente en todas y cada una de las áreas de investigación, como física, química, biología, computación, y lo que es de interés para este documento es el área de los mercados financieros [WP97].

Esta memoria está enfocada a abordar un problema financiero, relacionado con las series financieras de alta frecuencia y una forma particular de poder realizar pronósticos tomando en cuenta distintas características de naturaleza propia de este tipo de datos. Se pretende abordar esta problemática con metodologías computacionales, aplicando algunos conceptos matemáticos útiles para esta área. El problema es de carácter financiero, por lo que es necesario contextualizar el tema mediante conceptos, criterios y términos generales asociados al área.

Este capítulo tiene como objetivo introducir los conceptos de series de tiempo financieras, sus orígenes y la alta frecuencia.

1.1. Mercados Financieros

El mercado financiero es un espacio con marco institucional que permite poner en contacto a oferentes y demandantes para que efectúen transacciones financieras. La idea de mercado como foro organizado a la que acuden agentes económicos para efectuar transacciones queda reducida en el mundo financiero como las bolsas de valores [ME06].

El concepto de mercado financiero se utiliza en general para referirse a cualquier mercado organizado en el que se negocien instrumentos financieros de todo tipo, como acciones, divisas, etc. Dicho espacio para generar estas interacciones no necesariamente debe ser físico. Por otro lado el negociar instrumentos financieros implica a grandes rasgos: intercambiar instrumentos financieros, y definir su precio. Por ende, estos mercados están basados en las fuerzas de oferta y demanda, ubicando a todos los oferentes en el mismo lugar para facilitarle la búsqueda a los demandantes. Dentro de este tipo de mercado se distinguieron bloques de estudio en la economía moderna [JS84].

Una de las razones que hace importante este tipo de mercado, es su funcionalidad, ya que permiten por un lado aumentar el capital, siendo esto uno de los casos favorables, ya que también hay probabilidades considerables de disminuir el capital; comercio internacional, como en los mercados de divisas, por ejemplo Forex; y reunir a quienes necesitan recursos financieros, con los que tienen recursos financieros. Factores que permiten generar los efectos de oferta y demanda.

En este tipo de mercado se definen los siguientes términos [NSS03]:

- *Dealer*: Un dealer es un ente presente en los mercados que está dispuestos a comprar o vender.
- *Orders*: Operación de compra/venta de activos.
- *Bid price*: Precio al cual un *dealer* está dispuesto a comprar.
- *Ask price*: Precio al cual un *dealer* está dispuesto a vender.
- *Market orders*: instrucción del cliente al dealer, comprar o vender al mejor precio posible dentro de los valores actuales del mercado. Esto asegura la realización de la transacción, pero no el precio.
- *Limit orders*: es una orden para comprar a un valor máximo (precio determinado), o para vender a un valor mínimo (precio determinado). Esto le da al cliente el control

sobre el precio al que se ejecuta el comercio, sin embargo, no garantiza la realización de la transacción.

El conjunto de *Limit orders* forman los *books* para cierto activo, los cuales proveen información detallada de dicho instrumento. Con estos datos se forman los llamados bid-ask spreads, que es la diferencia entre el precios cotizados para una venta inmediata (oferta) y una compra inmediata (bid). También se generan los bid-ask quote, el cual define cotas para el precio de transacción.

1.1.1. Mercados bursátiles

Los mercados bursátiles están clasificados en los mercados de capitales, en donde se negocian activos financieros. Este tipo de mercado provee financiamiento por medio de la emisión de acciones y permiten luego el intercambio de estas. La aplicación más directa de este tipo de mercados, son las bolsas de valores, cuyo origen se remonta a finales del siglo XV en las ferias medievales de Europa. Las bolsas de valores se pueden definir como mercados organizados y especializados, en los que se pueden realizar transacciones de títulos de valores por medio de intermediarios autorizados. Estas bolsas ofrecen al público y a sus miembros facilidades, mecanismos e instrumentos técnicos que facilitan la negociación de títulos de valores susceptibles de ofertas públicas, a precios determinados mediante subasta [LZ98].

La principal función de las Bolsas de Valores es proporcionar a los participantes información objetiva, completa y permanente de los valores y las empresas inscritas en la bolsa, sus emisiones y las operaciones que en ella se realicen, supervisión de actividades. Las componentes de este sistema son los activos, instituciones financieras cuya misión es contactar demandantes y oferentes en los mercados donde se negocian los diferentes instrumentos o activos financieros.

Dentro de los estudios de la economía, se habla de que este tipo de mercado es de competencia perfecta, ya que posee características como: elevado número de compradores y vendedores; La decisión individual de cada uno de ellos ejercerá escasa influencia sobre el mercado global; Homogeneidad de los productos, es decir, no existen diferencias entre productos que venden los oferentes; Transparencia del mercado, todos los participantes tienen pleno conocimiento de las condiciones generales en que opera el mercado; Libertad de entrada y salida de empresas, todos los participantes, cuando lo deseen, podrán entrar o salir del mercado a costos nulos o casi nulos [Man11].

Sin embargo, Eugene Fama propuso la Hipótesis de eficiencia de los mercados [MF12], la cual dice que los mercados son eficientes cuando son capaces de trasladar a los precios de los instrumentos financieros, todos los datos relevantes, por lo tanto, el precio de esos activos refleja toda la información disponible, y lo hace de manera insesgada. Cuando se cumplen estas condiciones, el precio del instrumento financiero se comporta como un *Random Walk* [Fam65], por lo que los resultados no pueden ser predichos sistemáticamente.

1.2. Series de tiempo financiera

La mayoría de los fenómenos que se estudian temporalmente, deben tomar en cuenta la dinámica de los proceso con la finalidad tener una comprensión general del proceso. Una herramienta útil en dicho objetivo es el análisis de series de tiempo. Una serie de tiempo, es una secuencia de datos, medidos en determinado momento de tiempo y ordenados cronológicamente. Se pueden presentar casos de series de tiempo en una multitud de disciplinas como ingeniería, sociología, economía, finanzas por solo mencionar algunas de ellas.

El propósito fundamental es mostrar las técnicas que permitan hacer inferencias del proceso en estudio incluyendo su predicción. Esto se logra estableciendo modelos probabilísticos hipotéticos que representen a los datos; y en consecuencia, se lleva a cabo el proceso de ajuste que incluye desde la estimación hasta la predicción una vez que se determina un modelo satisfactorio. Los modelos de series deben considerar la naturaleza del fenómeno y determinar los factores que pueden ser incluidos en los modelos.

En particular, el análisis de una serie de tiempo financiera puede enfocarse en estudiar de forma teórica y práctica la valoración de cierto instrumento en el tiempo, como también los volúmenes transados, etc. Lo que se intenta es modelar la incertidumbre generada por las características propias de este fenómeno. Tanto la teoría económica y sus series temporales contienen un elemento de incertidumbre [Tsa05].

Las características diferenciadoras han propiciado numerosos trabajos en las áreas de econometría y economía financiera desde los años 60. Por ello, en el área financiera las evidencias sobre estructuras ha conducido a la formulación de distintos modelos matemáticos.

1.2.1. Descripción de serie financiera

Una tendencia en el estudio de las series financieras, es el estudio del precio de un instrumento en el tiempo. Existen varios tipos de precios que se pueden analizar, por un lado están los indicadores que se utilizaron en los primeros estudios formales, los cuales mostraban información al respecto de los precios de: *apertura*, *cierre*, *el más bajo*, *el más alto*, *promedio*. Esos datos lógicamente estaban orientados a tomar métricas respecto a la frecuencia *día*.

Una muestra gráfica de esta información queda como la siguiente:



Figura 1. Información de los precios de microsoft en la bolsa. Tomada desde <http://www.stockrageous.com/>.

Esos precios son a nivel diario, ya que los estudios en esos tiempos estaban orientados a esa frecuencia, como los flujos de caja, o análisis gráficos. Sin embargo la data estudiada actualmente corresponde a data *intra diaria*, la cual reside en los valores específicos en que se compró/vendió algún instrumento, ya sea acción, divisa, etc. Ese término se conoce como el *last price*, y corresponde al valor específico con el cual se transó el activo. La forma de como se realiza este evento es básicamente como el que se detalló anteriormente, es decir, mediante *market* o *limit orders*, donde existe una cola de compras/ventas obtenidas mediante los *limit orders* y que cuando se cumpla la condición asociada (de vender o comprar a cierto precio), se realiza la transacción y se registra dicho *last price*; y por otro lado, si se entra a comprar/vender directamente mediante una *market order* también se registra dicho valor. Juntando toda esa información, se puede generar una secuencia

temporal de los *last prices*. Estudios prácticos de estas series generalizan su gráfica con forma de *U*, [BHS12]. Por lo tanto, tomando en cuenta la alta frecuencia de estos datos, se torna muy complicado manejarlo de forma humana, es por esta razón que es necesario realizar análisis cuantitativos usando computadores.

A través de la historia, estas interacciones se realizaban personalmente en las bolsas de comercio, sin embargo con el avance de las tecnologías, se han implementado distintas plataformas de *trading electrónico*: un método de negociación de instrumentos financieros por medios electrónicos [Wes02]. Esta tecnología se utiliza para reunir a compradores y vendedores a través de la plataforma de comercio electrónico y las redes para crear lugares virtuales del mercado, como por ejemplo: National Association of Securities Dealers Automated Quotation (NASDAQ), New York Stock Exchange (NYSE), etc.

Las series financieras de alta frecuencia tienen características singulares en relación a otras series, como son:

- Frecuencia: la frecuencia en la observación de los datos es mayor que en otro tipo de series de tiempo, llegando incluso a unidades de fracciones de segundos.
- Heterocedasticidad: ocurre cuando la varianza de las perturbaciones no es constante a lo largo de las observaciones. Este factor hace inadecuados los modelos desarrollados para series estacionarias.
- No estacionalidad: en muchos casos la media y varianza se presentan como no estacionarias.
- Información oculta: se pueden encontrar distintos tipos de patrones a diferentes niveles de escala.

1.2.2. High frequency trading

Con el paso del tiempo y la implementación de distintos sistemas de trading electrónico, la frecuencia de la data creció de manera orbitante. En estos sistemas se define un *Tick*, como unidad atómica de información de un mercado financiero, particularmente sobre un activo determinado. Esta unidad especifica una gran cantidad de parámetros como la marca temporal, precio, cantidad transada, etc. Los datos de alta frecuencia son un conjunto de datos de reportes detallados sobre la actividad y movimiento efectuados sobre un activo. Se reúne la información de los ticks con cierto intervalo de tiempo, el cual para este tipo de series es variable, llegando incluso a tener ventanas de tiempo de fracciones

de segundos. *Wei Sun*, señala en [eS07] que en la literatura acuñan distintos términos para referirse a este tipo de datos.

Los sistemas informáticos que realizan la labor de facilitar el comercio de instrumentos financieros, son los llamados *Electronic communication networks* (ECN). Se crearon en 1998, año que fueron autorizados por la *Securities and Exchange Commission*. La SEC es una organización estadounidense creada en 1933 [Has04], y tiene la responsabilidad de velar por el cumplimiento de las leyes federales las bolsas de valores. Un caso emblemático, fue el 6 de mayo del 2010, fecha en la cual ocurrió el *Flash-Crash* [ALUG11], dio a lugar a una quiebra financiera estadounidense en el que el índice Dow Jones Industrial Average se desplomó un 9 %.

1.3. High Performance Computing

La capacidad de cómputo de los procesadores actuales ha incrementado de sobremedida, al igual que la cantidad de procesadores que tienen las tarjetas de video. Además, la naturaleza de los problemas que se están estudiando, como simulación de tsunamis, series de tiempo financiera, etc. Generan una gran cantidad de cómputo, por lo que es necesario optimizar y hacer más eficientes los algoritmos. Es por esto que nace el concepto de HPC que es el área de desarrollo que aplica computadores de alto rendimiento (clusters) a tareas que necesitan el procesamiento eficiente de métodos complejos o de grandes volúmenes de información. La idea de base consiste en dividir los problemas en subproblemas que se asignan a diferentes cores de un nodo o diferentes nodos de cálculo, coordinados por un nodo maestro mediante una red de alta velocidad.

1.3.1. Graphics Processing Unit

Las siglas GPU provienen de Graphics Processing Unit, o en español Unidad de procesamiento gráfico. Para efectos de hardware, la GPU funciona como coprocesador, pudiéndose utilizar de forma simultánea a la CPU y así aprovechar el potencial que puedan ofrecer ambas al mismo tiempo. Una GPU es un procesador diseñado para llevar a cabo cálculos necesarios implicados en la generación de gráficos, ya sea, para un video juego, o como para una aplicación que utilice gráficos en 2D, o 3D. Hoy en día las GPU son muy potentes y pueden incluso superar frecuencias de reloj de una CPU antigua.

Una GPU está formada por cientos de pequeños núcleos que trabajan juntos para procesar los datos de alguna aplicación. Esta arquitectura de procesamiento paralelo

masivo es la que proporciona al GPU su alta capacidad de cálculo. Existen numerosas aplicaciones aceleradas en la GPU que brindan una forma rápida de acceder a la computación de alto desempeño (High Performance Computing). Durante los últimos años se han desarrollado nuevas tecnologías y arquitecturas que permiten sacar mayor provecho a sus capacidades [Owe07].

El concepto implícito en todo este tema es el paralelismo, que es una forma de computación en la cual varios cálculos pueden realizarse simultáneamente, siempre y cuando no existan dependencias secuenciales entre cálculos. Basándose en "divide y vencerás", principio que busca dividir los problemas grandes, para obtener varios problemas pequeños, que son posteriormente solucionados en paralelo.

La evolución de las tarjetas gráficas ha venido acompañado de un gran crecimiento en el mundo de los videojuegos y las aplicaciones 3D, realizándose grandes producciones de chips gráficos por parte de grandes fabricantes, como NVIDIA, AMD (ex ATI). Los recientes desarrollos sobre GPU abarcan distintas áreas de la ciencia, como problemas astrofísicos (n-body simulation), modelamiento molecular, computación financiera, etc. [KWm10]. En los últimos años también han aparecido conjuntos de herramientas y compiladores que facilitan la programación de las GPUs, como por ejemplo, NVIDIA CUDA, que cuenta con la comunidad más activa hasta la fecha en programación de GPUs.

1.4. Motivación

Las características que presentan las series financieras de alta frecuencia son un problema recurrente y el estudio en esta área se torna a veces complejo. Uno de los enfoques de solución para este tipo de problema es realizar predicción mediante el uso de descomposición Wavelet Multi-escala [BMSR06], y los resultados son alentadores. Este enfoque se ha utilizado en distintas áreas de la ciencia, y en particular en series financieras. Más en particular Zhang ha propuesto un modelo neural-wavelet [ZD01], mezclando redes neuronales con wavelet multiescala, los cuales llegan a buenas aproximaciones pero no logran competir en eficiencia temporal.

Los *objetivos principales* son:

- Implementar un modelo de predicción Neural-Wavelet basado en el modelo de Zhang y analizar su comportamiento con series financieras de alta frecuencia.
- Optimizar dichos algoritmos usando CPU y GPU para aumentar su rendimiento.

Los *objetivos secundarios*:

- Comparar los resultados de forma empírica con datos reales de la bolsa.
- Encontrar una combinación de carga para CPU y GPU que permita hacer más eficiente los cálculos.

1.5. Organización de la Memoria

Esta memoria se organizará con el siguiente esquema:

- Capítulo 2: Estado del arte: se estudiarán y repasarán las técnicas relativas a .
- Capítulo 3: Descripción formal del problema: se formalizará el problema a resolver, indicando las características consideradas.
- Capítulo 4: Solución propuesta: se definirá una propuesta de solución, con la respectiva metodología de implementación.
- Capítulo 5: Estudio experimental: se implementará y testeará la solución propuesta, comparando sus resultados.
- Capítulo 6: Conclusiones: se realizarán conclusiones generales del trabajo realizado y se detallarán posibles trabajos futuros.

Capítulo 2

Estado del arte

En este capítulo se estudiarán las aplicaciones y estudios hechos en el área. Para ordenar esta sección, se realizarán divisiones respecto al tema de fondo de estudio (HFT), los métodos cuantitativos a usar (neural-wavelet) y finalmente el medio que pretende optimizar el rendimiento (GPU).

2.1. High Frequency Trading

2.2. Wavelet Multiscale

2.3. Artificial Neural Network

2.4. Graphics Processing Unit

Capítulo 3

Descripción del problema

3.1. Descripción histórica

3.2. Descripción y formalización a usar

Capítulo 4

Solución propuesta

4.1. Algoritmo y fundamentos teóricos

Capítulo 5

Estudio experimental

- 5.1. Selección de data (sector, frecuencia)
- 5.2. Parámetros del algoritmo
- 5.3. Validación
- 5.4. Speed up: multicore, gpu, híbrido

Capítulo 6

Conclusiones

Bibliografía

- [ALUG11] B. Arndt, M. Lutat, T. Uhle, and P. Gomber. High frequency trading. 2011.
- [BHS12] B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the paris bourse. *the Journal of Finance*, 50(5):1655–1689, 2012.
- [BMSR06] D. Benaouda, F. Murtagh, J.L. Starck, and O. Renaud. Wavelet-based nonlinear multiscale decomposition model for electricity load forecasting. *Neurocomputing*, 70(1):139–154, 2006.
- [eS07] W. ei Sun. Quantitative methods in high-frequency financial econometrics: Modeling univariate and multivariate time series. 2007.
- [Fam65] E.F. Fama. Random walks in stock market prices. *Financial Analysts Journal*, pages 55–59, 1965.
- [Has04] J. Hasbrouck. Economic and statistical perspectives on the dynamics of trade in securities markets. *Teaching Notes, Stern School of Business, New York University*, 2004.
- [JS84] M. Jensen and C. Smith. The theory of corporate finance: a historical overview. 1984.
- [Kir07] D. Kirk. Nvidia cuda software and gpu parallel computing architecture. In *International Symposium on Memory Management: Proceedings of the 6 th international symposium on Memory management*, volume 21, pages 103–104, 2007.
- [KWm10] D.B. Kirk and W.H. Wen-mei. *Programming massively parallel processors: a hands-on approach*. Morgan Kaufmann, 2010.
- [LZ98] R. Levine and S. Zervos. Stock markets, banks, and economic growth. *American economic review*, pages 537–558, 1998.
- [Man11] N.G. Mankiw. *Principles of economics*. South-Western Pub, 2011.
- [ME06] F.S. Mishkin and S.G. Eakins. *Financial markets and institutions*. Pearson Education India, 2006.

- [MF12] B.G. Malkiel and E.F. Fama. Efficient capital markets: A review of theory and empirical work*. *The journal of Finance*, 25(2):383–417, 2012.
- [NSS03] Y. Nevmyvaka, K. Sycara, and D.J. Seppi. Electronic market making: Initial investigation. In *the Proceedings of Third International Workshop on Computational Intelligence in Economics and Finance*, 2003.
- [Owe07] J. Owens. Gpu architecture overview. In *ACM SIGGRAPH*, volume 1, pages 5–9, 2007.
- [Tsa05] R.S. Tsay. *Analysis of financial time series*, volume 543. Wiley-Interscience, 2005.
- [Wes02] J.P. Weston. Electronic communication networks and liquidity on the nasdaq. *Journal of Financial Services Research*, 22(1):125–139, 2002.
- [WP97] T.J. Watsham and K. Parramore. *Quantitative methods in finance*. Cengage Learning Business Press, 1997.
- [ZD01] B.L. Zhang and Z.Y. Dong. An adaptive neural-wavelet model for short term load forecasting. *Electric power systems research*, 59(2):121–129, 2001.