***High Frequency Trading* (HFT) – O que é?**

*High Frequency Trading* - HFT, é um método de realizar transações baseada em algoritmos que são executados em alta frequência; no mercado financeiro, é utilizada como forma de transacionar ativos, tudo o que pode ser negociado no mercado tais como ações, câmbio, opções e etc, de forma rápida, gerando um grande fluxo de compra e venda de ativos que possuem alto grau de liquidez, o quão rápido e fácil é possível comprar ou vender um ativo, algo necessário para que seja possível negociá-los em alta velocidade, ou enviar ordens de mercado, tanto para a sua execução, quanto como parte de uma estratégia de manipulação dos valores dos ativos, algo bem comum neste tipo de operações. (hft for dumies)

Para ser considerado uma operação de HFT é necessário que se cumpram quatro princípios básicos definidos no livro (hft for dumies), que são:

Ser executada por algoritmos que operam em alta velocidade: Com isso não há um grande ganho por execução, na verdade os ganhos são bem pequenos devido à natureza do hft, que tem como objetivo o ganho por número de execuções, acumulando ganhos de centavos ou até mesmo frações deles milhares de vezes gerando um valor substancial no fim do período determinado para a operação.

Uso de hardware de alto desempenho localizado próximo ao local de negociação dos ativos: São realizadas milhares de operações por dia que geram pequenos ganhos por vez, então o melhor método de maximizar os ganhos é realizar o maior número de transações possível e garantir a compra e venda do ativo no valor desejado, logo a latência da rede exerce grande influência sobre as transações, e como há um limite na velocidade que os dados podem trafegar na rede, diminuir fisicamente a distância que eles precisam percorrer é um método de minimizar o tempo de comunicação e melhorar o desempenho dos algoritmos.

Uso de tipos específicos de ordens: São ordens que tem bem definido, pelos algoritmos que as executam, a forma que serão dispostas no mercado, como será mostrada na lista de ordens e as influências que ela sofrerá de acordo com as variações que ocorrem na lista.

Grande fluxo de ordens sendo transmitidas: operações baseadas em hft geram um grande fluxo de ordens, seja para serem efetivadas ou como forma de avaliar o mercado e perceber possíveis pontos onde se pode obter ganho.

O hft trouxe muitos benefícios as operações de mercado, já que a sua utilização permitiu uma redução nas taxas das transações, e gera uma grande liquidez no mercado por fazer milhares de transações diárias o que é muito benéfico para os investidores e para as empresas que disponibilizam ativos no mercado. (hft for dumies)

**O que hft faz?**

O hft não faz nada sozinho, é apenas uma forma de executar algoritmos de forma rápida e com o menor tempo de comunicação possível, a sua utilização consiste em estratégias de operação de mercado que podem ser divididas em quatro classes definidas por (aldridge 2013 2013 hft) que são:

1. Arbitragem
2. *Directional event-based trading*
3. *Automated market making*
4. *Liquidity detection*

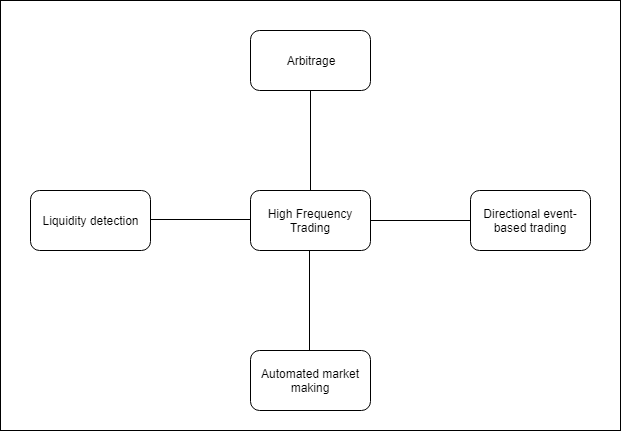


Figura 1 Classes de estratégias hft (hft aldridge 2013)

*Arbitrage*, o foco do trabalho, haverá um tópico especifico sobre ele.

*Directional event-based trading*: é uma classe de estratégias baseadas em predições estatísticas sobre tendências que os ativos apresentam no curto prazo, variação dos valores dos ativos em determinado período, impacto de eventos sobre os ativos, como uma noticia afeta o valor de mercado do ativo (hft aldridge 2013).

*Automated market making:* Ele substitui os *broker-dealers*, pessoas que realizavam a compra e venda dos ativos para os clientes nas *exchanges*, locais onde são negociados os ativos, por processos automáticos executados por algoritmos, ele compreende as formas tradicionais de negociação no mercado baseado em dois tipos *inventory-driven* e *information-driven* tendo como objetivo minimizar o risco que se tem ao negociar, dessa forma é possível escolher com que npivel de agressividade, quão grande é o risco aceitável nas operações, se quer operar (hft aldridge 2013).

*Liquidity detection:* São estratégias baseadas em intervenções de mercado efetuadas através de manipulação de ordens enviadas ao mercado gerando variações nos valores dos ativos e gerando pequenos ganhos por vez.

**Mercados onde o hft pode ser utilizado**

Para que um mercado esteja apto á aplicação de estratégias de operação por hft, é necessário que ele tenha 2 características fundamentais, como define (aldridge 2010, hft), que são: alta liquidez e possibilidade de execução por algoritmos, que, em conjunto com as normas de regulamentação dos tipos de ativos negociados são as características principais que avaliam a compatibilidade do ativo para o uso de hft nas transações (aldridge 2010, hft). A figura 1.2 mostra a frequência com que cada tipo de ativo deve ser negociado de acordo com a sua liquidez.

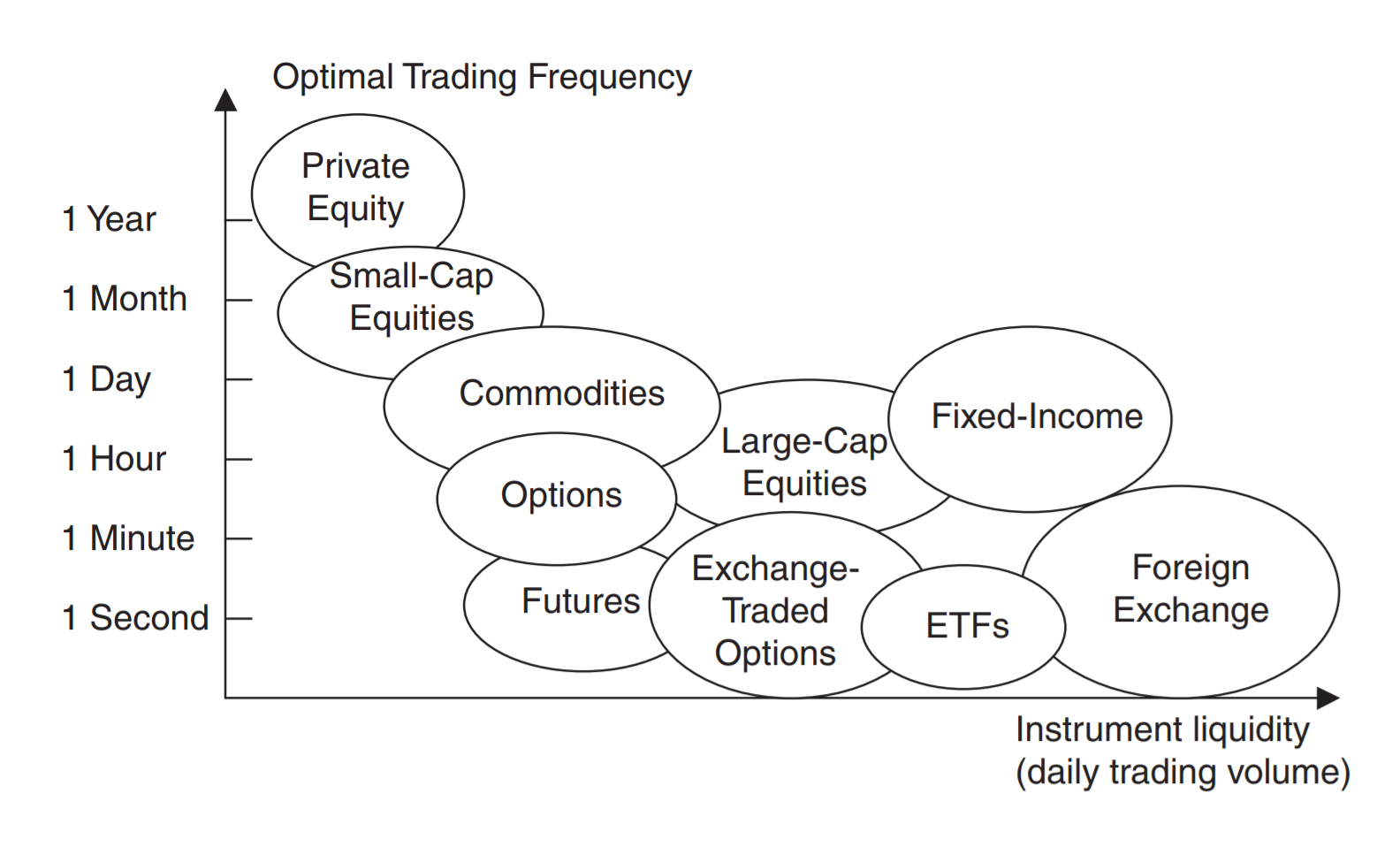


Figura 2 frequência de transação de cada tipo de ativo de acordo com a sua liquidez. (aldridge 2010, hft)

Para uma utilização ótima de hft para negociar os ativos quanto mais abaixo e a direita do gráfico melhor será a compatibilidade pois, a liquidez será maior e a frequência de transações é elevada, a partir desta simples análise do gráfico é possível ver que o mercado de *Foreing exchange* é o melhor para a utilização de hft, e com toda razão, este mercado movimentava 5.1 trilhões de dólares por dia em abril de 2016, dados obtidos do *bank of international settlements*, tendo alta liquidez e tempos de transação mínimos, entretanto o mercado de *commodities* não parece ter uma boa compatibilidade já que possui pouca liquidez e a frequência de transações é baixa, na ordem de 1 transação por hora ou até mesmo por dia, o que pode ser verificado pela forma como as *commodities* são negociadas, geralmente por contratos do tipo *spot*, que são semelhantes a ir ao mercado, compra-se um produto e o recebe fisicamente, tornando inviável a utilização de hft (aldridge 2010, hft).

**Vantagens do hft**

Segundo (aldridge 2010, hft) há duas grandes vantagens em utilizar hft que são;

Falta de concorrência: nos novos sistemas de transações eletrônicas a utilização de hft gera grandes ganhos pela simples falta de concorrência.

Não é uma soma zero: com diversas estratégias possíveis é praticamente garantido que todos os operadores tem a possibilidade de obter algum ganho utilizando seu próprio modelo de negócio.

Além dessas duas vantagens também é possível citar os impactos positivos no mercado que a utilização do hft produz, tais como, o aumento da liquidez dos ativos e a diminuição no custo por operação (hft for dumies).

**Arbitragem o que é?**

Arbitragem é uma das estratégias de operação por hft, sendo a mais popular entre elas, ela consiste em aproveitar as diferenças nos preços dos ativos e gerar ganho a partir disso, a um custo de operação zero (zhaodong, hft), por exemplo: o ativo X é vendido na *exchange* Y por um valor Z, e está sendo comprado pela exchange W por um valor Z+i, como o custo das operações de compra e venda é zero a diferença i nos valores de compra e venda do ativo representa o ganho obtido durante a transação.

Em seu livro (zhaodong, hft), define dois tipos de arbitragem, a determinística e a estatística, a primeira também conhecida por ser *risk-free*, não apresenta risco de perda na transação, é o tipo menos comum de arbitragem, manipulando ativos que devido a sua natureza não apresentam desvalorização ao decorrer do tempo, como o ouro por exemplo, logo não importa quando irá ser feita a venda do ativo, é garantido que se irá ter um ganho sobre ele. Arbitragem estatística é a mais utilizada pelos *traders*, ela possui a mesma base da arbitragem determinística, porém utiliza analises estatísticas, fórmulas matemáticas que apontam para uma posição de ganho com uma probabilidade X de acontecer, quanto maior X melhor, geralmente são determinados valores mínimos para X como uma forma de proteção, se X baixar do valor limite as negociações são encerradas, evitando grandes perdas, para realizar as negociações, nesta abordagem são negociados 2 ou mais ativos, há a possibilidade de operar com apenas uma ativo mas é uma atividade menos comum, que possuem algum tipo de correlação, apresentam tendências ascendentes e descendentes semelhantes, a figura 3 exemplifica dois ativos correlacionados, é possível ver claramente como ambos os ativos seguem a mesma tendência de variação até que um deles sai de sincronia, mostrando uma boa oportunidade para obter ganho, porém volta a convergir após algum tempo, mostrando uma forte correlação entre os ativos.

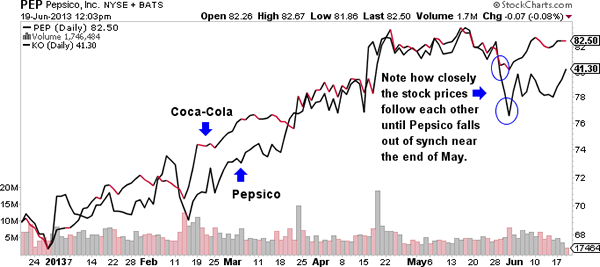


Figura – 3 Ativos correlacionados, fonte: Nasdaq

Por negociar diversos tipos de ativos, a abordagem estatística apresenta em suma quatro grandes riscos em suas operações, que são:

*Exchange rate*: taxa de variação do valor do ativo, esta taxa pode variar gerando uma situação de soma zero, não há ganho com a venda dos ativos, ou resultar em perda, o ativo de posse inicia um movimento de queda enquanto o ativo relacionado inicia um movimento de alta, por mais que esse risco represente um grande problema, as variações de valor no mercado são de forma lenta, salvo alguns casos excepcionais que podem ser previstos, sendo um risco considerado baixo (zhaodong, hft).

Não convergências dos valores: em estratégias que se baseiam no ganho pela compra e venda de ativos correlacionados, quando estes divergem da média para seus valores, vende-se – ativo que está em alta e compra-se o que está em queda, o ganho surge quando os valores voltam a convergir para a média, o ativo vendido inicia um movimento de queda e o comprado um movimento de alta, gerando ganho em ambos os lados, porém há o risco dos valores não convergirem novamente gerando uma grande perda pela aquisição do ativo que está em queda, este é um risco considerado crítico e geralmente ocorre quando a empresa dona do ativo vai a falência ou quando há uma onda crescente de inflação que desvaloriza uma moeda adquirida (zhaodong, hft).

Contratos de ativos são de tamanhos diferentes: geralmente ocorre quando são negociados ativos físicos, como o ouro por exemplo, em mercados diferentes, as unidades de medida dos mercados podem ser diferentes, como kg e lb, logo há uma diferença na quantidade que será negociada podendo resultar em perdas, se o contrato em um mercado por a venda do ouro por kg e os contratos de compra em outro mercado for em lb, havera divergência de 100g por kg negociado, compra-se 1kg porém só é possível vender 900g, resultando em uma perda, apesar de parecer critico uma perda de 10% ao realizar várias transações pode ser possível equalizar o valor, a cada 9 compras de ouro é possível fazer 10 vendas, resultando em perda zero, tornando o risco baixo, quase nulo (zhaodong, hft).

Falta: para obter ganho deve-se tentar adquirir ambos os ativos em seu valor ideal de compra e venda, porém pode ocorrer de se obter um dos ativos pelo valor desejado e o outro não, gerando a falta, este é um erro crítico e de maior ocorrência, para tentar contorná-lo há duas abordagens, vender o ativo adquirido e não obter ganho porém não ter perdas, ou, adquirir o outro ativo a um valor maior porém próximo do ideal gerando em um ganho menor (zhaodong, hft).

Apesar dos riscos apresentados a arbitragem estatística é de certa forma segura, pois os riscos podem ser mitigados com abordagens de controle para evitar grandes perdas, cada estratégia possui chances diferentes de cair em um dos casos de risco acima devido ao tipo de ativo que elas negociam, contudo em suas implementações isso já é considerado e medidas de proteção são implementadas.

**Estratégias da arbitragem estatística:**

A arbitragem estatística possui diversas estratégias que podem ser aplicadas de acordo com o tipo de ativo que se deseja negociar, isso abre oportunidades para que todo tipo de *trader* possa obter ganho no mercado, as principais estratégias são ilustradas na figura 4.

Para o trabalho serão abordadas as estratégias que tem mais influência de poder computacional do que da análise estatística, já que estas apresentam melhorias mais perceptíveis quando utilizadas em conjunto a hardware de alto desempenho.

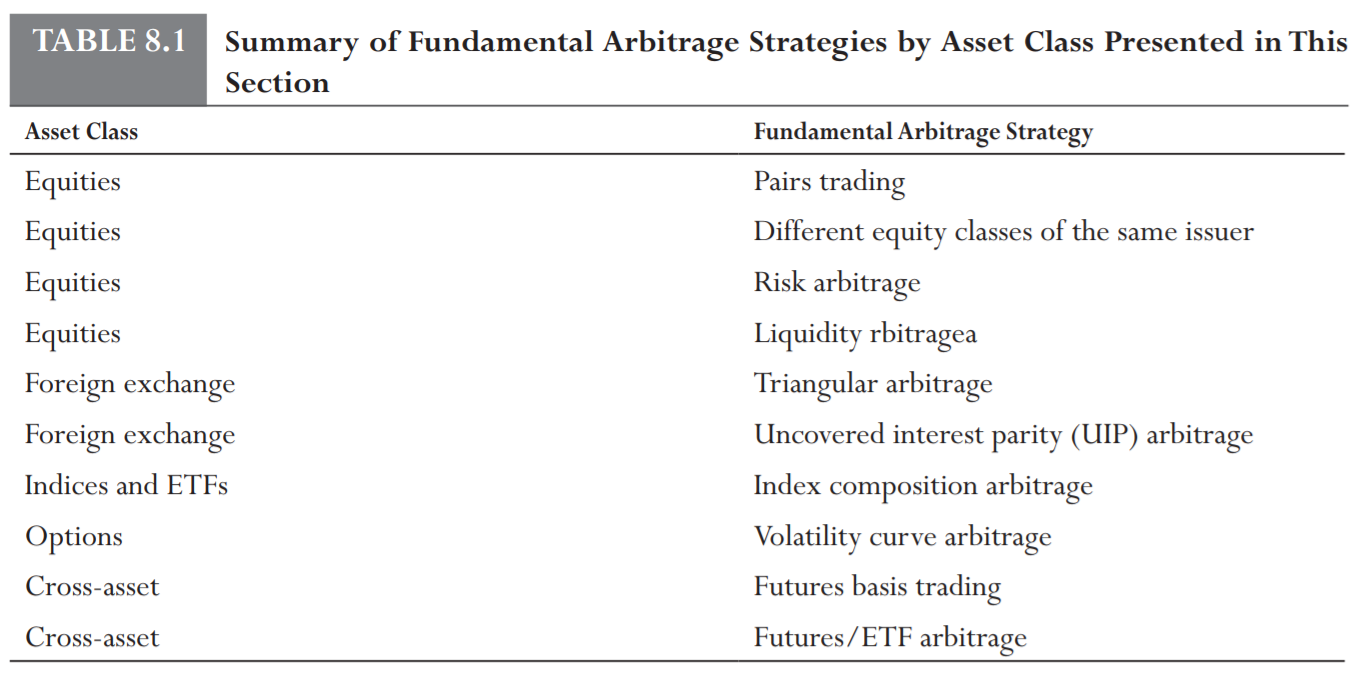


Figura – 4 Estratégias de arbitragem estatística (Aldridge 2013).

Pair trading: é a estratégia mais simples e utilizada pelos traders, ela é baseada na negociação de ativos correlacionados, como os mostrados na figura 3; o método de escolha para os ativos que serão negociados consiste em analisar uma quantidade de pelo menos 30 amostras por dia para realizar a inferência sobre o ativo, ou até 500 amostras diárias para uma inferência forte, em um período considerável de tempo, em torno de 1 a 2 anos, (Aldridge 2013), tendo obtido os dados calcula-se a o quadrado da diferença entre os valores dos ativos, para o par de ativos que possui a menor diferença, é calculada a média para a diferença dos valores da amostra E(), e o desvio padrão σ() para um dos ativos que será negociado, depois basta avaliar os valores na inequação:

Onde: é o valor do ativo no tempo t

é o valor do ativo no tempo t

Se verdadeira deve-se vender o ativo i e comprar o ativo j, caso contrário se:

Compra-se o ativo i e vende-se o ativo j.

Quando os valores convergirem para a média, representando um ganho satisfatório, fecha-se as posições, caso a movimentação dos ativos comecem a divergir muito a política de gerenciamento de risco é ativada evitando grandes perdas (Aldridge 2013).

Forex Exchange: Se trata da negociação de câmbio, o método mais simples de arbitragem forex é a arbitragem triangular, ela se aproveita da diferença que pode haver durante a conversão de câmbio entre moedas (Aldridge 2013), o exemplo a seguir ilustra bem o seu funcionamento.

Serão negociadas 3 moedas que podem ser da mesma instituição ou de instituições diferentes, o importante é que a taxa de conversão entre elas favoreça a utilização de arbitragem, as moedas serão USD, EUR e CAD onde as taxas de conversão de cada par são, dados reais coletados dia 07/06/2018 as 21:00 horário de Brasília:

USD/EUR = 0,84761; EURO/HKD = 9,25760; HKD/USD = 0,127448

Com essas taxas e iniciando com um investimento de 1000 usd seguindo arbitragem triangular obteremos:

1000 USD \* 0,84761 = 847,61 EURO

847,61 EURO \* 9,25760 = 7846,834 HKD

7846,834 HKD \* 0,127448 = 1000,06 USD

1000,06 – 1000 = 0,06 USD

Desta forma obtemos um ganho de 0,06 USD, parece um ganho ínfimo, e realmente é, entretanto esse ganho é instantâneo ao final de cada transação, ou seja a cada vez que esta conversão for feita se terá um ganho de 0,06 USD, considerando as taxas de câmbio acima, e ele estará disponível imediatamente após o fechamento da operação, então se considerarmos uma média de 500 transações por dia, número baixo considerando uma aplicação que opera em hft, teremos um ganho de 30 USD ao final do dia, 3% ao dia de ganho bruto, um rendimento altíssimo se considerado que o investimento com maior rendimento no brasil em 2017 foi o ibovespa com 26,86% segundo o site g1.globo.com, porém qual a probabilidade que a taxa de cambio se mantenha exatamente neste valor durante o dia todo, praticamente zero, notícias e acontecimentos fazem as taxas de câmbio oscilarem constantemente, contudo moedas fortes apresentam pouca variação em suas taxas como pode ser visto nos gráficos a seguir:



Figura – 5 Taxa de conversão para as últimas 12h USD para EURO: Fonte XE Corp.



Figura – 6 Taxa de conversão para as últimas 12h EUR para HKD: Fonte XE Corp.



Figura – 7 Taxa de conversão para as últimas 12h HKD para USD: Fonte XE Corp.

Como pode ser notado ao analisar os gráficos em um mês as taxas de conversão tem variações de décimos de centavo, é perfeitamente possível que essas variações façam com que não seja vantajoso a realização da arbitragem, vamos supor que em 90% dos casos essa variação inviabilize a operação, o ganho diário de 30 USD passara a ser de 3 USD, 0,3% o que ainda é bem superior aos investimentos mais comuns.

**Hardware:**

Sistemas para *trading* em hft necessitam de hardware de alto desempenho para que possam funcionar, atualmente no mercado há três grandes tipos de hardware que podem ser utilizados para rodar os sistemas e são eles, unidade de processamento central CPU, unidades de processamento gráfico GPU e as *field programable gate arrays* FPGA, cada uma delas possui um nível de desempenho e um custo atrelado, permitindo maior liberdade de escolha de acordo com a necessidade do sistema, entretanto para o dimensionamento correto do hardware para o sistema é preciso entende-los para se fazer a escolha correta(aldridge 2013). A figura 1 mostra a evolução de desempenho e o custo de cada unidade de hardaware.

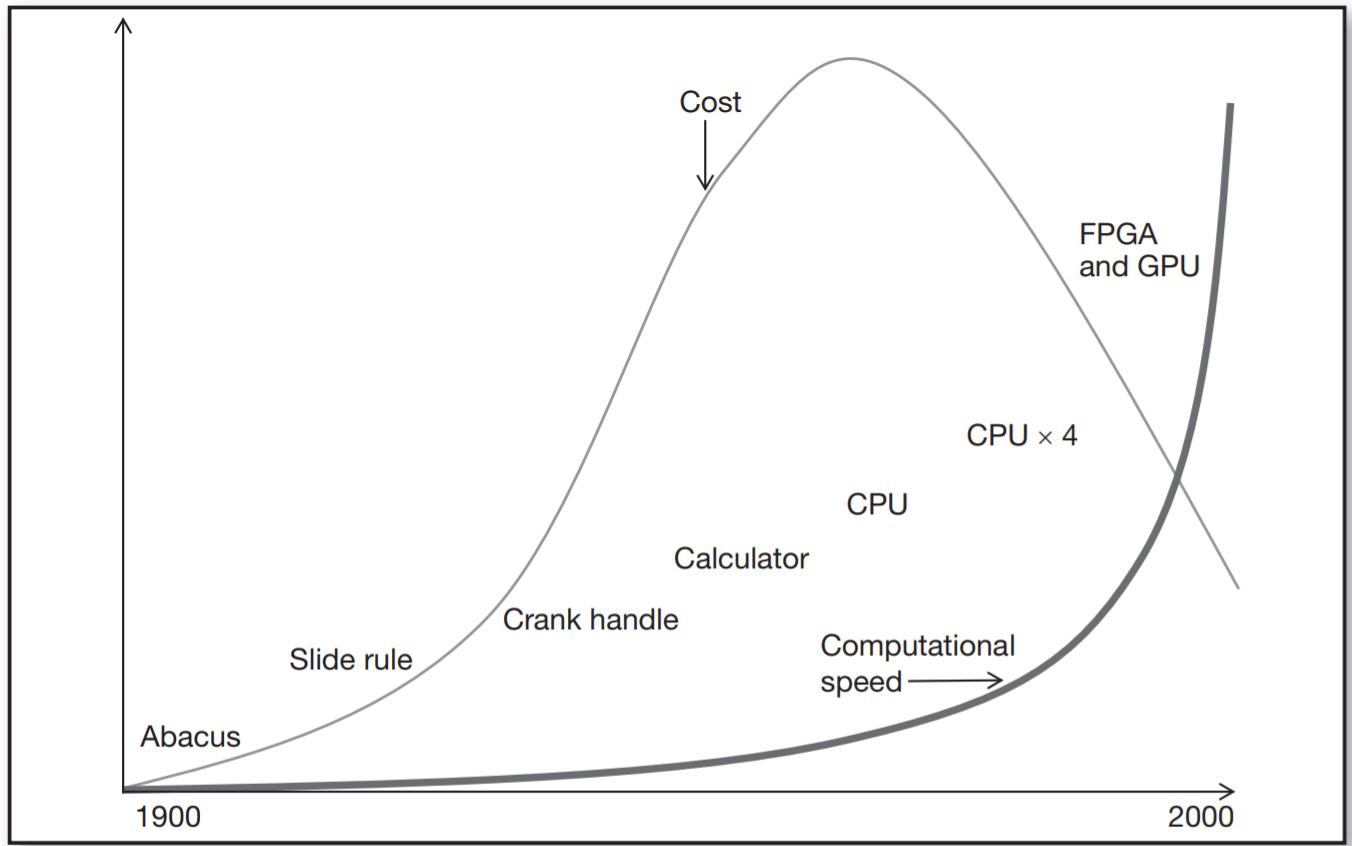


Figura – 1 Custo e poder computacional dos sistemas de hardware (aldridge 2013).

CPU – é a mais comum e utilizada em todo tipo de computador pessoal, ela é composta basicamente da unidade de processamento logico aritmética ALU, módulos de memória, registradores e cache, e unidades de escalonamento, ela pode ser *singlecore* ou *multicore*, os chips mais avançados usam o padrão *multicore*, que possuem algumas otimizações, como memória interna compartilhada e caches de alto desempenho, apesar de tudo isso, esta é a opção com menor desempenho entre as três citadas, pois, precisa atender a diversos tipos de sistemas perdendo desempenho em troca de flexibilidade de uso, e também com um baixo custo, no mercado brasileiro é possível encontrar chips *multicore* de alto desempenho entre 1000 e 2000 reais (Aldridge 2013). A figura a seguir ilustra as partes que compõem uma CPU multicore.

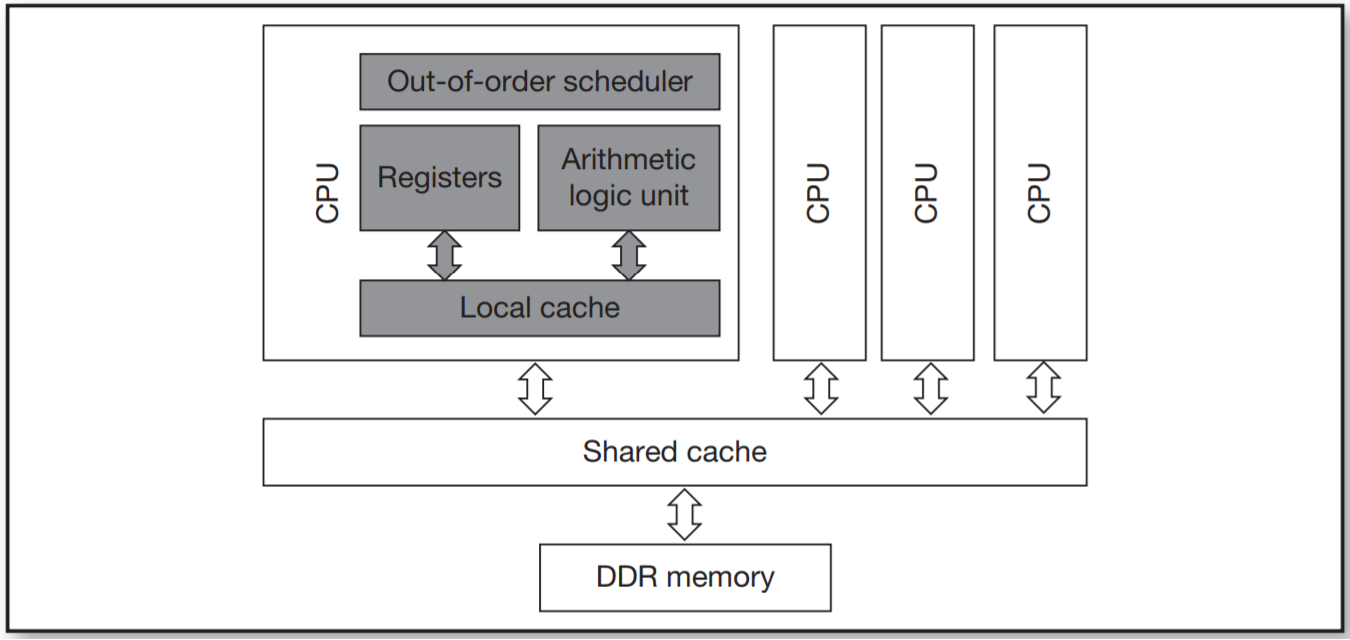


Figura – 2 Componentes de uma CPU (Aldridge 2013).

GPU – são chips de processamento gráfico que diferentemente de uma CPU possuem mais unidades logicas de aritmética ALU, por serem chips de propósito específico eles ganham em desempenho perdendo a flexibilidade e dispensando componentes que não são necessários, abrindo espaço para mais módulos ALU, quando comparados a uma CPU, que funcionam de forma paralela, gerando um alto desempenho, o que as torna ótimas para a realização de cálculos, por possuir unidades que operam em paralelo a programação de aplicações para este tipo de chip requer maior experiência por parte do programador, qualquer um pode escrever um código e executá-lo em uma GPU, contudo, sem o controle de paralelismo, o desempenho será muito semelhante ao de uma CPU desperdiçando grande parte do poder computacional. O custo de uma GPU no mercado brasileiro é mais elevado em comparação com uma CPU, unidades de alto desempenho custam em torno de 2000 reais e podem chegar a 3000 reais, o interessante é que quando comparamos os modelos top de linha tanto de CPU quanto de GPU, os valores parecem convergir girando em torno de 4000 a 5000 reais para ambas, o que torna as GPU muito mais atraentes por apresentarem um custo benefício superior (Aldridge 2013). A figura a seguir ilustra as partes que compõem uma GPU.

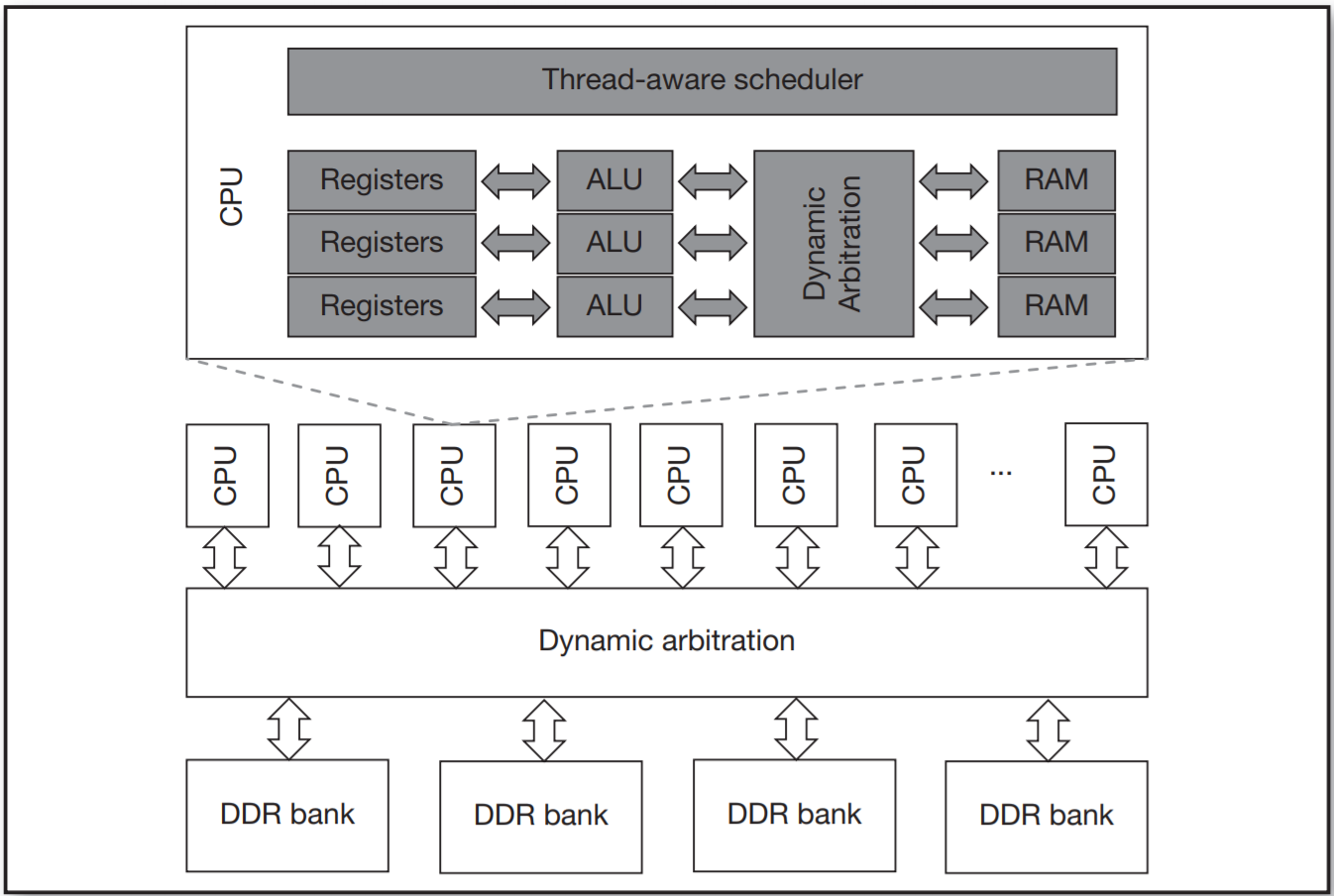


Figura – 3 componentes de uma GPU (Aldridge 2013).

FPGA – é o que há de melhor em questão de desempenho, eles são compostos por módulos de memória, interpretadores de sinais, e principalmente por unidades lógicas, que podem ser programadas via linguagem descrição de hardware, VHDL ou Verilog, a figura 4 mostra um exemplo de código em VHDL, a codificação não é tão simples quanto criar um programa em linguagem tradicional, ela exige maior experiência por parte do programador e um conhecimento mais profundo sobre o funcionamento de hardware, utilização de portas lógicas, manipulação de registradores e paralelismo são essenciais, pois são a base para o desenvolvimento, e permitem criar um hardware específico para aplicação que operam com paralelismo puro, ou é possível utilizar processadores desenvolvidos para elas, Nios e Leon são exemplos, e utilizar instruções customizadas que aumentam o desempenho do processador, não tem o mesmo desempenho de um hardware completamente dedicado, o algoritmo é criado em linguagem C, por exemplo, e é executado sobre o processador com as otimizações, mas é muito superior a uma CPU (Aldridge 2013). Ter um hardware específico pode não parecer muito vantajoso, GPUs apresentam bom desempenho e requerem menos esforço para desenvolvimento, entretanto ter instruções próprias para a aplicação reduzem o tempo de processamento pois são capazes de realizar operações em menos ciclos de *clock*, uma instrução que consome 5 ciclos de clock em uma CPU pode ser feita em 1 ou 2 ciclos utilizando hardware específico, se bem implementados algoritmos inteiros podem ser executados em poucos ciclos de clock devido à natureza paralela que os FPGAs possuem (Aldridge 2013). O custo de um FPGA é extremamente abrangente, pode-se encontrar placas com boa capacidade por 85 dólares, até placas especificas que chegam facilmente a 5000 dólares, o que torna difícil estabelecer um custo benefício, porém elas são, com certeza as indicadas para aplicações de alto desempenho. A figura 5 ilustra as partes que compõem um FPGA.

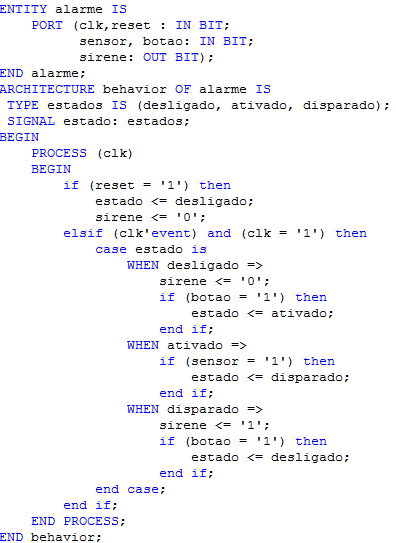


Figura – 4 Máquina de estados em VHDL, slides de aulas Prof Dr Vanderlei Bonato

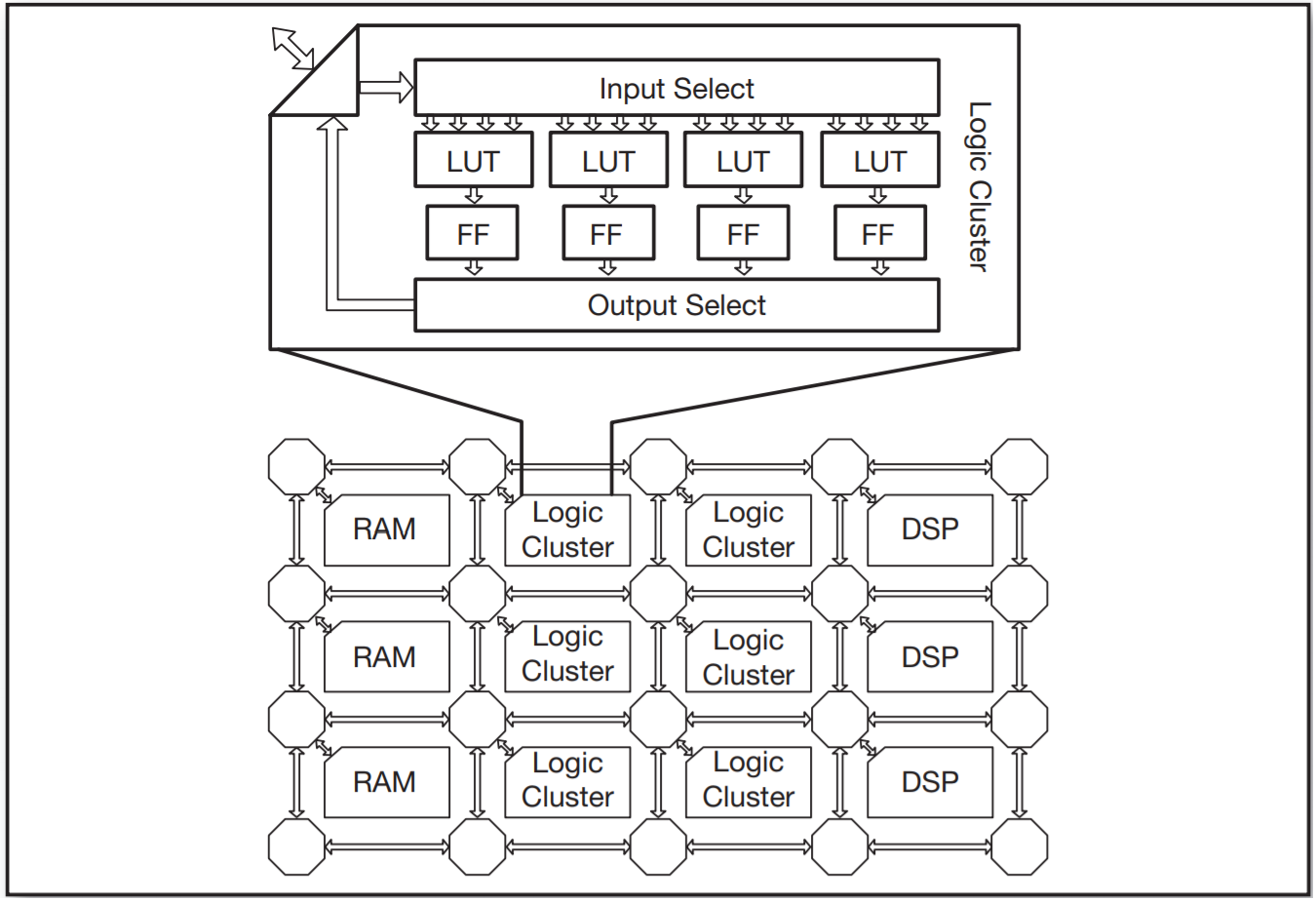


Figura – 5 Componentes de um FPGA (Aldridge 2013).

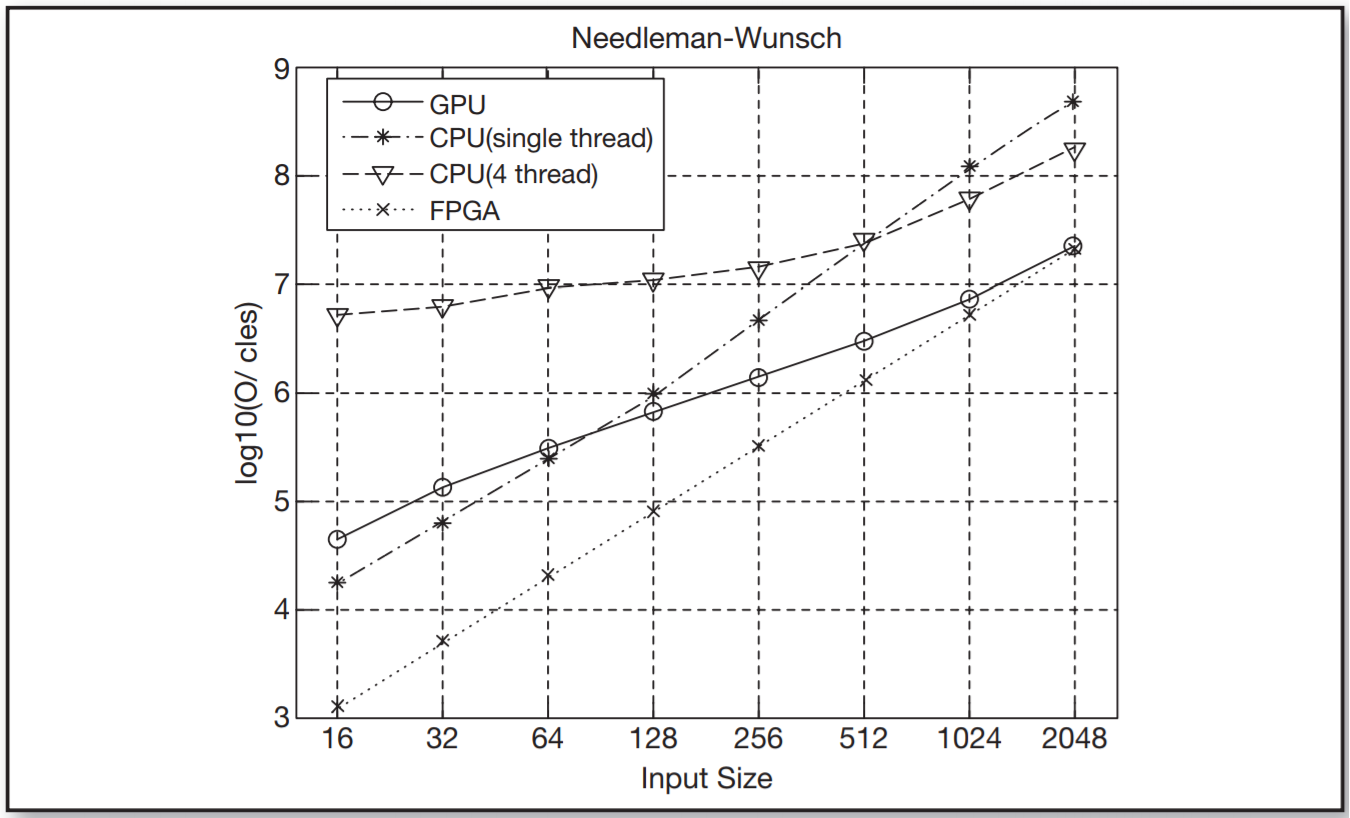


Figura – 6 Comparativo de performance entre FPGA, GPU, CPU, CPU *multicore* (Aldridge 2013).

Como é possível ver na figura 5, compara a quantidade de ciclos de *clock* necessários para processar um número x de variáveis distintas, FPGAs são a melhor opção quando se busca desempenho, principalmente para aplicações que tratam de poucas variáveis, arbitragem triangular geralmente fica nessa área de até 2000 entradas e como o algoritmo é pequeno, é possível utilizar unidades mais simples de FPGA obtendo um excelente custo benefício, quando o número de variáveis aumenta consideravelmente o desempenho de FPGA e GPU começa a convergir tornando as GPUs opções mais atraentes para utilização, estratégias que precisam analisar grandes quantidades de dados podem ter um melhor custo benefício se implementas sobre GPUs, dito isto a escolha do hardware que será utilizado na implementação depende não só do desempenho puro que ele pode apresentar e sim de uma combinação entre, custo, desempenho e necessidade da estratégia que será utilizada.

Além do hardware, a comunicação durante as operações de hft tem papel crucial na performance do sistema, sendo o ponto onde geralmente se encontram os gargalos dos sistemas, devido a latência de rede, velocidades de barramento e etc.

**Protocolos de comunicação:**

A comunicação entre os processos é composta por 3 níveis de protocolos, UDP, TCP-IP, FIX/ITCH/OUCH, cada um é utilizado para tratar um tipo especifico de dado, de acordo com as necessidades que eles apresentam (Aldridge 2013). A figura 6 ilustra a complexidade dos protocolos utilizados durante a comunicação, que é proporcional as necessidades de cada tipo de dado que é transmitido, dados mais sensíveis necessitam de maior cuidado aumentando a complexidade do protocolo.

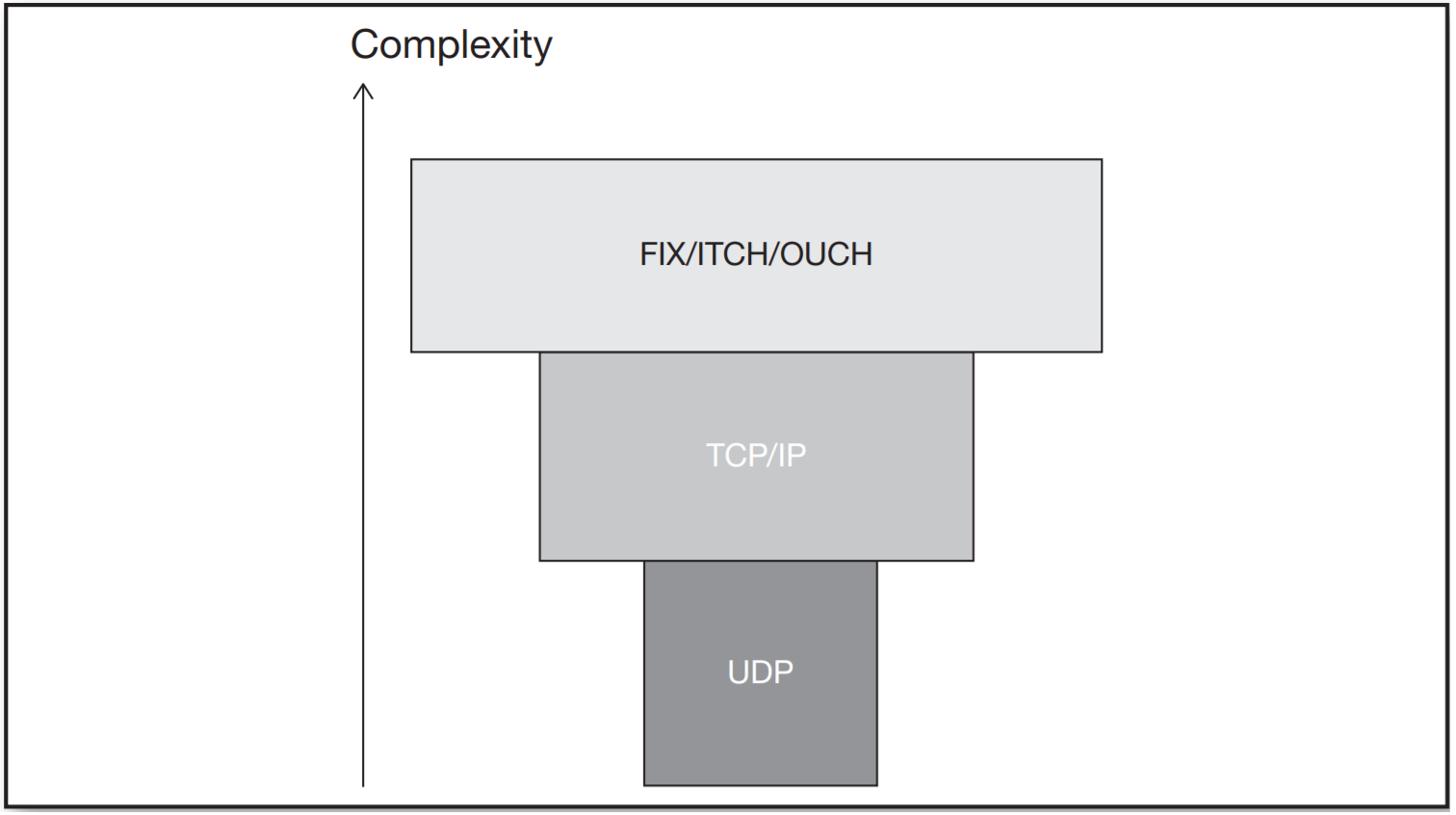


Figura 6 – complexidade dos protocolos de comunicação (Aldrigde 2013).

UDP - é o mais simples e por isso o mais rápido, entretanto ele não garante que os dados foram entregues, sendo geralmente utilizado para transmissão de informações de grande volume e que não são sensíveis a perda de dados, tais como as informações sobre os valores que os ativos de interesse foram negociados, como é uma grade quantidade de dados transmitida e a perda de uma informação não é relevante, outro dado será recebido em breve com a nova atualização para a informação perdida (Aldridge, 2013); contudo deve-se prestar atenção sobre a integridade dos dados recebidos, pois eles serão utilizados nas estratégias de negociação, e se os valores obtidos apresentares um desvio considerável dos valores reais a estratégia não terá validade, pois não irá gerar os ganhos previstos e há grande possibilidade de gerar prejuízos, o mesmo vale para qualquer previsão sobre o comportamento do ativo que seja baseada nos valores obtidos, as previsões geradas pelos modelos podem não representar o comportamento real que o ativo terá, indicado para uma posição divergente da real tornando a previsão inutilizável. (Aldridge, 2013)

TCP/IP – é um protocolo de transmissão de dados mais seguro e que garante o recebimento e a integridade dos dados obtidos, sendo utilizado para transmitir informações mais sensíveis e que necessitam de confirmação tais como ordens de execução, devido ao tamanho maior dos pacotes, possuem mais dados que permitem verificar a integridade dos dados, e a necessidade de confirmação de recebimento dos pacotes a comunicação é mais lenta, em torno de 1/3, que a UDP (Aldridge, 2013). A imagem abaixo demonstra a utilização dos pacotes de acordo com as suas características.

(FIX - *Financial Information exchange*), ITCH, OUCH – são protocolos de padronização das informações transmitidas, para que sejam facilmente entendidas pelo computador, são utilizados em conjunto com os protocolos tcp/ip e udp. Os protocolos ITCH e OUCH, são mais eficientes por transmitirem os dados em binário, não necessita de processamento para transformar os dados de texto para binário e vice versa (Aldridge, 2013), porém são protocolos proprietários Nasdaq, diferente do protocolo FIX, que é gratuito e baseado no formato xml, que mesmo sendo mais lento é o mais comum, devido a isso será o protocolo abordado, ambos são utilizados para as operações que necessitam transmissão de dados, valores dos ativos, ordens, transações (Aldridge, 2013).

As mensagens no formato FIX, são escritas com uma tag numérica que indica o significado do dado, e são separadas por um SOH, s*tart of header*. (Aldridge, 2013) a seguir um exemplo de mensagem no formato FIX.

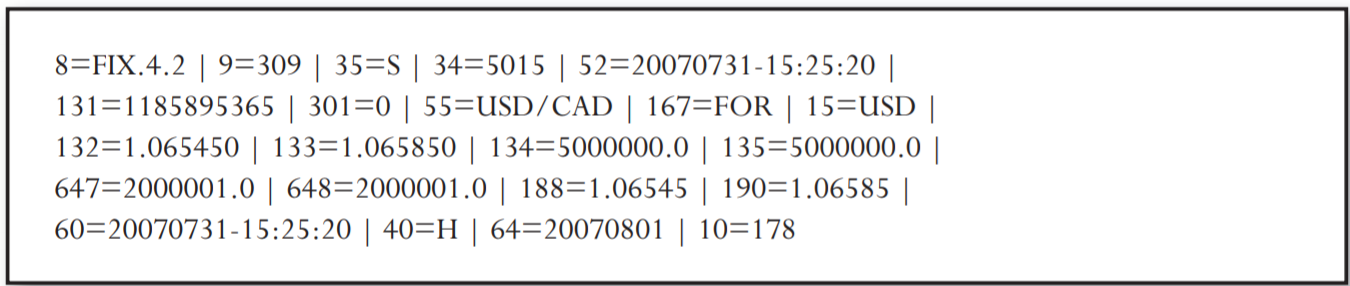


Figura 7 – exemplo de mensagem FIX.

As informações passadas pela mensagem são:

Versão do protocolo, tag 8;

*Time stamp*, tag 52;

Identificador do ativo “USD/CAD”, tag 55;

Tipo do ativo, FOR, de *foreign exchange*, tag 167;

Moeda base, “USD”, tag 15;

Melhores valores de compra e venda, tags 132 e 133;

Quantidade disponível para compra e venda, tags 134 e 135.

A figura 8 ilustra a utilização dos protocolos de acordo com o tipo de dado, os mais sensíveis utilizam o protocolo TCP/IP enquanto os demais utilizam o protocolo UDP, e em ambos os casos o protocolo FIX é utilizado para formatar a mensagem de forma a facilitar a comunicação.

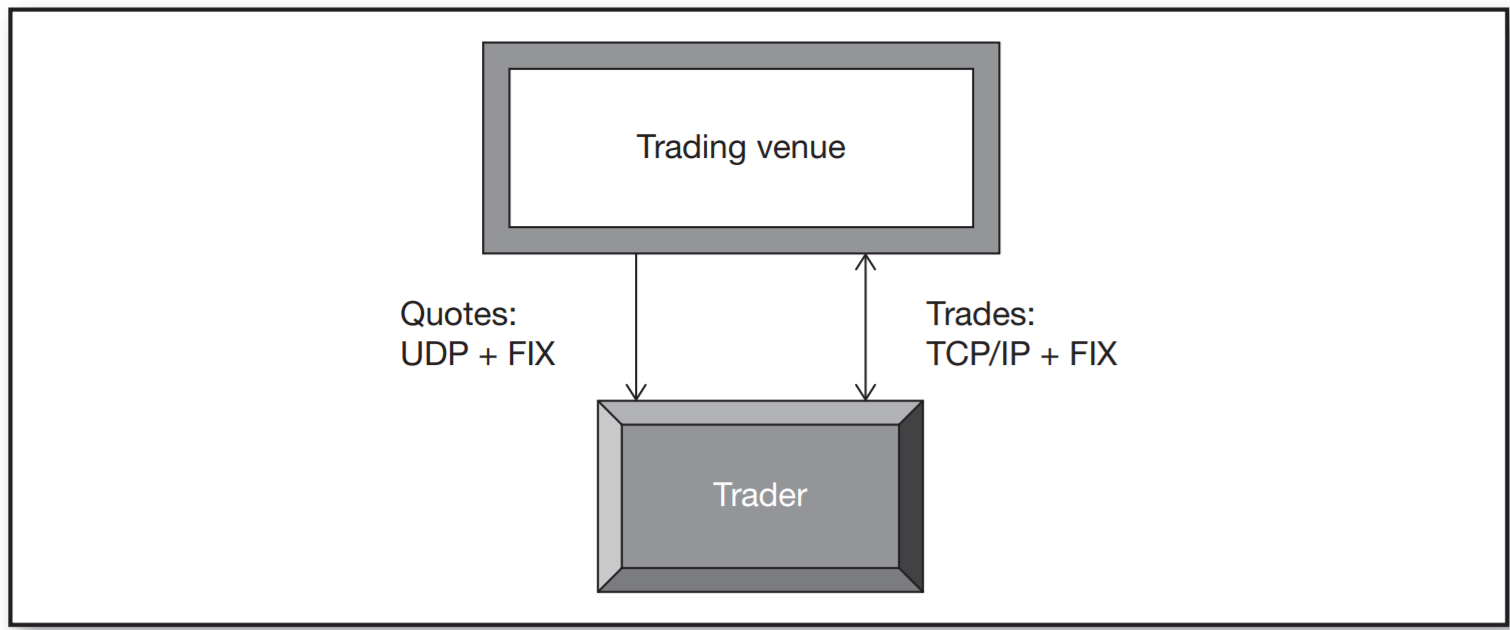


Figura 8 – utilização dos protocolos de comunicação.

**Modelo base das comunicações:**

Independente do protocolo utilizado as mensagens transmitidas seguem uma arquite-tura base comum apresentada por (Aldridge, 2013) na imagem a seguir.

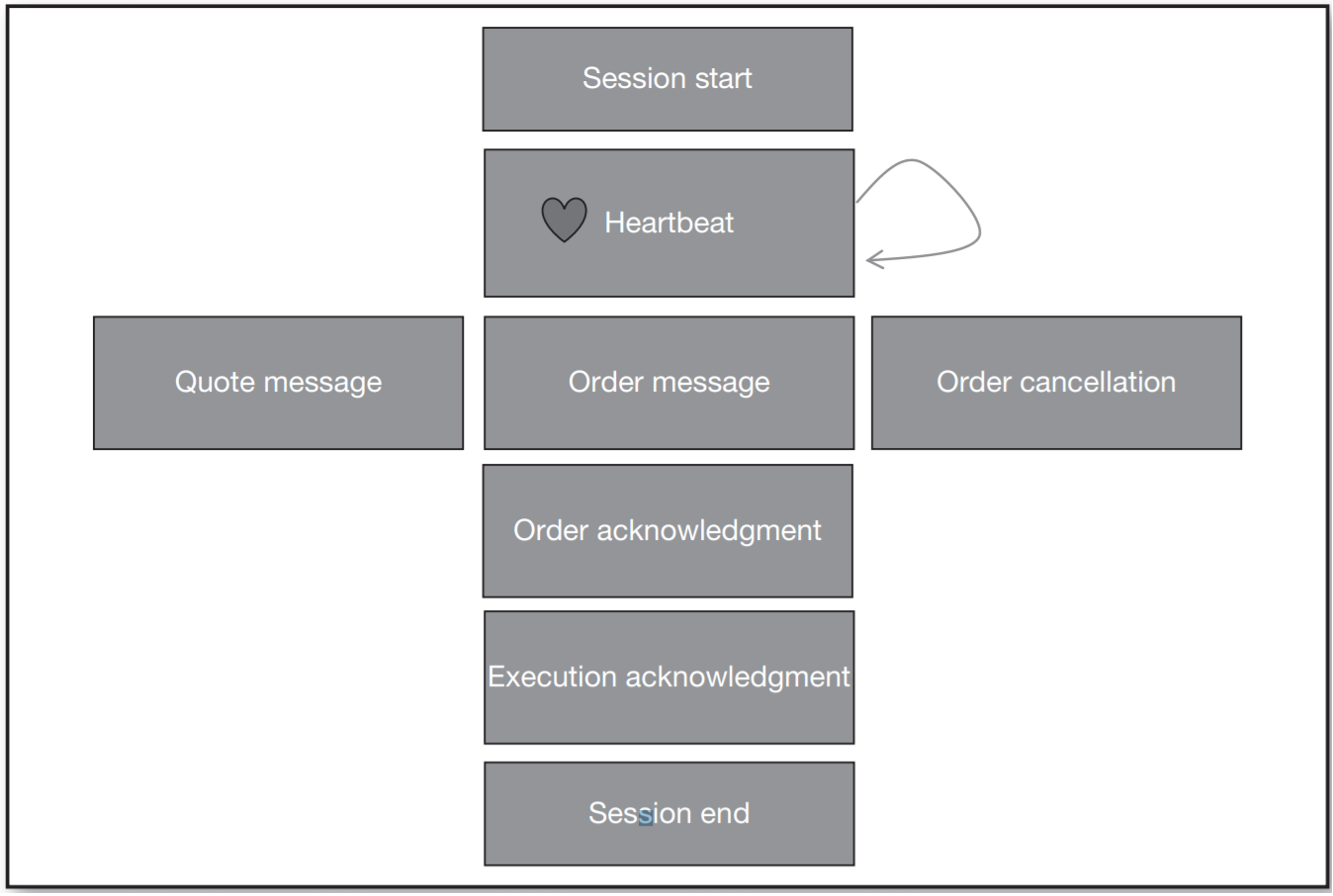


Figura – 9 Os elementos principais das mensagens definidos por (Aldridge, 2013) são:

***Session Start*:** a mensagem que inicia a comunicação, ela notifica que a entidade está aberta a transações e deseja abrir um canal de comunicação.

***Heartbeat*:** mensagem que notifica que a entidade está ativa e que apta a realizar transações, se a entidade falhar em enviar um número predeterminado de heratbeats, as entidades que possuem conexão com ela fecham o canal de comunicação, necessitando de uma nova *session start* para reiniciar a comunicação.

***Quote message,* informação sobre valor do ativo:** mensagens que informam os valores de compra e venda dos ativos sendo negociados.

**Mensagens sobre ordens:** informações sobre os tipos das ordens, tamanho e execução.

**Cancelamento de ordem:** mensagem para o cancelamento de ordens.

**Confirmação de ordem:** confirmação de que a ordem foi executada ou cancelada.

**Confirmação de execução:** dados referentes a execução tais como, horário, valores e quantidades negociadas.

***Session end*:** informa que a entidade encerrou as operações.

**Eficiência:**

A eficiência da comunicação será medida em dois quesitos que são de extrema importância para as aplicações que utilizam hft, que são a velocidade e a segurança na transmissão de dados, já que quanto mais rápido a comunicação a quantidade de transações por unidade de tempo aumenta representando, idealmente, maior ganho, e como os dados transmitidos são sensíveis, informações sobre a compra e venda de ativos e sobre o executor, a segurança é imprescindível a aplicação.

Com relação a segurança dos dados, tanto o protocolo TCP/IP quanto o UDP, não oferecem criptografia enviando os dados em texto plano e permitindo sua leitura de forma fácil uma vez interceptada a mensagem, o protocolo FIX permite que os dados sejam criptografados, porém há um custo para tal, que dependendo do tipo de transação pode ser alto demais e não valer a pena; com isso em mente é importante escolher o tipo de rede de comunicação com a *exchange* correto para o tipo de transação desejado, dessa forma pode-se fixar um nível de segurança mínimo desejado e ter o máximo de desempenho possível com relação a velocidade de transmissão dos dados (Aldridge 2013). Para realizar a comunicação há três tipos distintos de arquitetura de redes que podem ser utilizados, Cliente-Servidor, *Peer-to-Peer* e *co-location*, cada uma com um nível de segurança e desempenho diferente da outra, além, é claro, de custos diferentes que são proporcionais ás capacidades da rede. A figura 10 ilustra cada tipo de arquitetura de rede e suas capacidades de transmissão de dados definidas por (Aldridge, 2013).

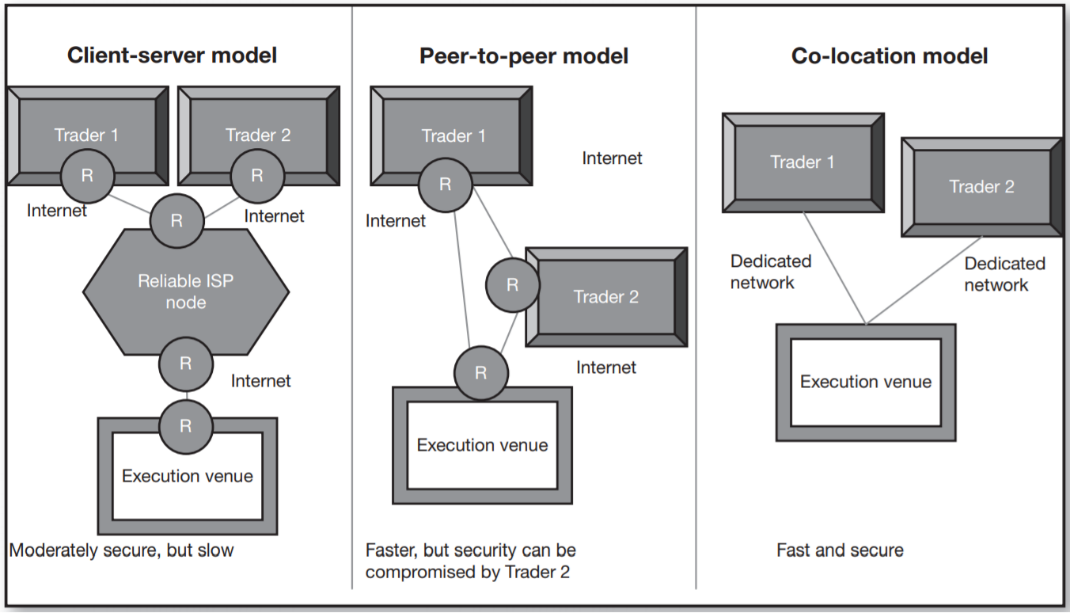


Figura 10 – Arquiteturas de rede e suas capacidades.

Cliente-servidor: a mais utilizada, toda a comunicação é feita através de nós que fazem a ligação física entre as entidades que desejam trocar informações, os nós ISP *Internet service providers*, que são gerenciados por grandes empresas de comunicação, possuem medidas de segurança que impedem, até certo ponto, que as mensagens que trafegam pela rede não serão espionadas por terceiros, garantindo um certo nível de segurança para os dados que trafegam na rede, entretanto são poucos os nós que gerenciam o tráfego, em torno de 80, e dependendo do fluxo de mensagens, pode haver congestionamento na rede, diminuindo a velocidade que os dados trafegam, fazendo com que este tipo de rede tenha como características principais, um nível de segurança razoável e uma velocidade de tráfego de dados lenta(Aldridge 2013).

*Peer-to-Peer*: é uma arquitetura descentralizada, não há nós centrais que realizam a comunicação entre as entidades, os dados são transmitidos através da rede formada pelos usuários conectados a ela, cada usuário opera como um nó de comunicação, devido a forma que é implementada ela possibilita velocidade superior a cliente-servidor, já que não há congestionamento na rede, todavia, não há nada que impeça que a mensagem seja lida pelos usuários que servem como intermediadores para a comunicação, tornando essa rede menos segura, mesmo que a randomização da distribuição dos pacotes que contêm a mensagem seja suficiente para que um nó não receba a mensagem por completo, dessa forma as características principais desta arquitetura são: velocidade superior a cliente-servidor e falta de garantia de segurança dos dados (Aldridge 2013).

Co-location: a melhor e a mais cara, trata-se de um link de comunicação direta com o servidor da *exchange*, os computadores das entidades são posicionados fisicamente próximos a *exchange*, e possuem uma conexão por cabos direta com os servidores, dessa forma a rede é completamente segura contra espionagem de dados, pois não possui intermediadores para realizar a conexão, além de velocidade de trafego altíssima, depende apenas da velocidade com que os dados podem fluir no cabo, tanto que a distância física entre as entidades e os servidores é regulamentada, já que alguns metros de diferença podem geram vantagem para quem está mais próximo, no entanto todos estes benefícios tem um custo, é necessário ter um local físico próximo a exchange para guardar os computadores que farão a comunicação, o que eleva muito o custo deste tipo de rede de comunicação, dessa forma as características que representam este tipo de rede são: alto nível de segurança, alta velocidade de transmissão de dados e custo elevado em relação as outras arquiteturas (Aldridge 2013).

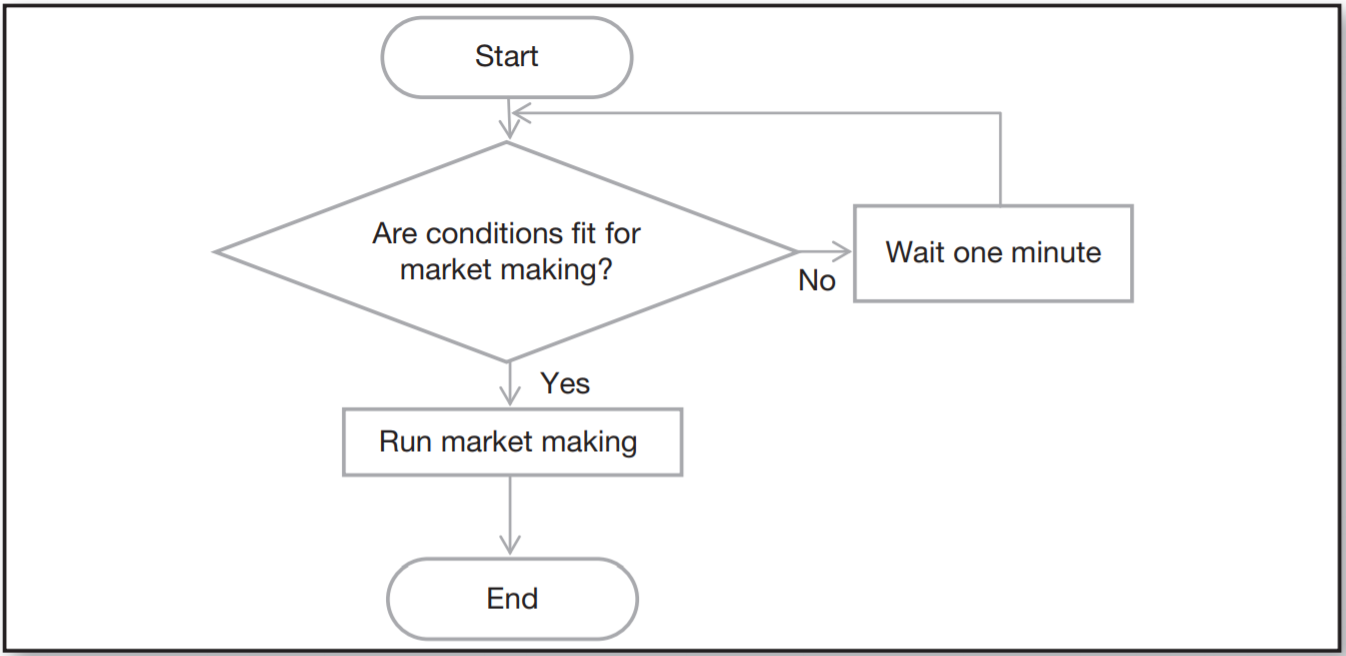
Analisando os tipos de arquitetura de rede pode-se escolher o melhor tipo para cada aplicação, levando em consideração o foco do trabalho em arbitragem, a utilização de uma rede tipo cliente servidor com protocolo udp, e dados em texto plano, é uma boa opção para aquisição dos valores dos ativos negociados, por serem dados de visualização pública, e a utilização de protocolo TCP/IP com mensagens criptografadas para a realização das ordens, o tempo gasto para criptografar a mensagem não é tão relevante considerando a latência da rede, é uma boa opção para realizar as transações.

**Implementação:**

A implementação do sistema de transações com hft se dá de forma semelhante a de qualquer sistema computacional, é um algoritmo como qualquer outro, que, por especificação do sistema, utiliza os protocolos de comunicação descritos para realizar a troca de mensagens e segue as diretivas para ser considerado de alta frequência.

O algoritmo, idealmente, seria implementado em hardware, utilizando um FPGA e uma linguagem de descrição de hardware, VHDL ou Verilog, o que permite o maior desempenho possível, porém é perfeitamente possível a codificação dos algoritmos em linguagens tradicionais, tais como, C, C++ ou Java, este método perde em desempenho porém permite uma maior portabilidade do sistema, já que o programa poderá ser executado independente de um hardware especifico, e também traz uma maior facilidade de codificação, implementar um algoritmo em Java é muito mais simples do que implementar diretamente em hardware, o que pode ser um atrativo, pois permite que qualquer pessoa com um mínimo de conhecimento de programação possa implementar o algoritmo (Aldridge 2013).

Com relação a lógica de desenvolvimento, ela é bem simples na verdade, são poucos os passos que devem ser cumpridos para a codificação do algoritmo, ilustrado por (Aldridge 2013) na figura 10.

Figura – 10 Estrutura do algoritmo

A implementação é inicializada verificando então as condições do mercado, se forem satisfatórias inicia-se a negociação, caso não espera-se um determinado tempo e avalia-se novamente as condições, este passo é repetido até que se encontre uma posição favorável a iniciar as negociações, o tempo de espera e as condições de aceitação são dependentes da estratégia de negociação adotada, após a negociação estar concluída a execução é encerrada (Aldridge 2013). O código com a estratégia é geralmente bem pequeno, como sugere o modelo acima, algo em torno de 50 linhas, ele apenas implementa a fórmula inerente ao tipo de estratégia adotada, o que dá a impressão errada de ser simples, apesar do código com a estratégia ser pequeno, somente ele não é suficiente para um bom sistema de negociação por hft, são necessários códigos para captar e tratar as informações dos ativos que se pretende negociar, o que inclui comunicação entre processos e utilização de protocolos de comunicação que não são triviais, e políticas de gerenciamento de risco, que podem ser consideradas como o ponto mais importante para qualquer sistema, são elas que protegem o investimento em caso de variações bruscas de mercado ou caso a movimentação do ativo entre em uma situação de perda, somados, esses códigos podem atingir facilmente a casa de 60.000 linhas, mostrando que são sistemas complexos e que deve se tomar cuidado para não ter a ideia errada e acabar criando um sistema que só irá gerar perdas (Aldridge 2013).