International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 08 | Aug 2023 www.irjet.net p-ISSN: 2395-0072

DATA SCIENCE APPROACH TO STOCK MARKET ANALYSIS

Sneh Patil[1], Yash Bharat Jadhav[2], Neel Gude[3], Anish Shivram Gawde[4], Mohammed Naif[5], Aditi Bombe[6]

Abstract - *Stock analysis is a technique that aids buyers and sellers for investors and traders. Investors and traders attempt to obtain an advantage in the markets by making educated decisions by researching and assessing both historical and recent data. We only care about the information that is pertinent to us in the information-overflowing world. The material can be categorised according to a number of disciplines, including engineering, the arts, science, history, sports, geography, and economics. In these circumstances, it is crucial for us to focus on the information that is pertinent to us.[16] The information that is provided is poorly organised and only provides information on the current situation, particularly when it comes to the sector of finance. On the stock market, virtually all significant economic transactions take place at a dynamic rate known as the stock value based on market equilibrium. A precise forecast of this stock trend in advance could result in enormous rewards.*

Key Words: **Data Science, Stock Market, Algorithms, Trade, Data Analysis**

1.INTRODUCTION

The stock market is quite complicated and prone to volatility, and it is influenced by a number of factors. Due to the market's unpredictability, investors can benefit or lose money. Stock analysis is a tool used by traders and investors to assist them decide what to purchase and sell. Investors and traders strive to obtain an advantage in the markets by making educated decisions by researching and analysing both historical and recent data. A precise forecast of this stock trend in advance could result in enormous rewards. [14, 21] Two strategies are used in this paper to achieve the prediction. Utilising different technical indicators is one of them. These technical indicators are used to predict upcoming shifts in stock trend. [21] The technical indicators used in this stock market forecasting tool will substantially support the predictability.

The second strategy, which is based on the Hidden Markov Model, is more probabilistic. This model has been broad and is more suited for dynamic systems. used when solving pattern recognition issues. [22, 23] It is focused on statistical techniques and has a robust capacity for handling new data. Based on historical datasets that complement current stock price behaviour, this model interpolates the two datasets using suitable neighbouring price elements.[16, 15] As a result, a prediction is made regarding the variable of interest's stock trend tomorrow. [22] Fundamental analysis, which considers the financial performance of the specific firm we are analysing as well as significant financial news regarding the company, can be used to analyse the stock market. [14] This leads us to news analysis, where we must parse through all the pertinent company news and papers to obtain the necessary data so that we can decide whether or not to invest. The "Theory of Demand and Supply" underlies the functioning of the stock market [12, 18]. It asserts that supply and demand are always negatively correlated.[18] As a result, as demand rises, purchasers seize control of the stock market, which then turns bullish. On the other side, as soon as supply enters the market, sellers seize control and stock prices begin to decline sharply. The "Head and Shoulder Theory" [18], which states that the highest price of any stock is considered the head and the next peak high is called the shoulder, can also be used to analyse the stock market. The majority of the indicators we'll encounter in this study have a statistical basis. As a result, each technical indicator's design step involves some assumptions.

2. UNDERSTANDING STOCK MARKET

Although practically anything can now be purchased online, each commodity often has its own market. For instance, individuals travel to farms and the outskirts of cities to acquire Christmas trees, go to the neighbourhood timber market to buy wood and other supplies for house renovations, and shop at Walmart for their weekly groceries. Such a focused marketplace acts as a venue for various buyers and sellers to connect, communicate, and conduct business. One can be sure that the pricing is fair because there are so many market participants. For instance, if there is only one Christmas tree vendor in the entire city, he will be free to charge whatever he wants because there will be nowhere else for customers to go. If there are many tree sellers in a single marketplace, they will have to compete with one another to draw customers to purchase their goods. It will be a fair market with transparent pricing because purchasers will have a wide range of options. Even when buying online, customers compare prices from several merchants on the same app or across different apps or websites to find the greatest offers, which compels the different online vendors to provide them the best deal. A stock market is a comparable venue set aside for the trading of various securities in a regulated, secure, and controllable setting. The stock market ensures fair pricing practises and transparency in transactions by bringing together hundreds of thousands of market players who want to purchase and sell shares of various companies. Modern computer-aided stock markets operate electronically, which makes them more and more advanced. Earlier stock markets used to

work on the issue and transact in paper-based physical share certificates.

In short, stock markets offer traders a safe, regulated environment where they can confidently trade shares and other permissible financial products with little to no operational risk. The stock markets function as both primary and secondary markets while adhering to the regulator's established guidelines. The stock market, which serves as a primary market, enables businesses to issue and sell their shares to the general public for the first time through the initial public offering (IPO) procedure. This practise aids businesses in obtaining the funding they want from investors. It basically implies that a business divides itself into a certain number of shares, such as 20 million, and sells a certain portion of those shares, such as 5 million, to the general public at a certain price, such as $10 a share. A corporation needs a market where these shares can be sold to make this process easier. The stock market provides this marketplace. If everything goes according to plan, the corporation will sell the 5 million shares at a price of $10 each and make a profit of $50 million. In anticipation of growing share prices and any prospective income in the form of dividend payments, investors will receive business shares that they can anticipate to hold for the term of their choice. The corporation and its financial partners pay the stock market a fee for its services as a facilitator of this capital-raising process.

The stock exchange also acts as the trading platform that enables routine buying and selling of the listed shares after the first-time share issuance IPO exercise, also known as the listing process. The secondary market is comprised of this. Every trade that takes place on the stock exchange's platform during secondary market activity generates a charge for the stock exchange. The task of guaranteeing price transparency, liquidity, price discovery, and fair dealings in such trading activities falls to the stock exchange. The exchange maintains trading systems that effectively manage the buy and sell orders from diverse market participants because nearly all major stock markets across the world now operate electronically. In order to facilitate transaction execution at a price that is fair to both buyers and sellers, they carry out the price matching function. A listed firm may later conduct other offerings, such as follow-on offers or rights issues, to issue new, extra shares. They might even buy back shares or take them off the market. Such trades are facilitated by the stock exchange. The S&P 500 index and the Nasdaq 100 index, which provide a gauge to follow the movement of the whole market, are only two examples of the numerous market-level and sector-specific indicators that the stock exchange frequently develops and maintains.

The stochastic oscillator and stochastic momentum index are other techniques. All corporate news, announcements, and financial reporting are also maintained by the stock exchanges and are typically accessible on their sites. Various additional corporate-level, transaction-related activities are also supported by a stock market. Profitable businesses, for instance, may reward investors by providing dividends, which are often derived from a portion of the company's profits. All of this information is maintained by the exchange, which also has the potential to facilitate some of its processing.

2.1 Interval

Every second, there are millions of trades made in the market. So, in order to analyse the data, we must categorise these trades according to the time and period in which they occurred. These trades can be divided into short-term and long-term time frames. One minute, two, five, ten, fifteen, thirty, and sixty minutes are additional categories for intra-day. Long-term intervals fall within the daily, weekly, monthly, and so forth categories. Count the Costs Any stock will be associated with 4 different price categories at every interval. High, Low, Open, and Close prices are all displayed. The greatest value at which it was exchanged during that period of time is what is meant by the high price. The bargain price represents the lowest amount it went for during that time. The price at which the stock was last traded during that interval is known as the "close price," while the "open price" refers to the first deal that occurred during that period.

2.2 Trend

Any stock's demand at any one time may be higher than its supply or lower than its available supply, depending on the trend. As a result, we can divide this Trend into two categories, namely Bearish and Bullish. When there is more demand than supply at a given period, a stock is said to be in a bullish trend. A stock is said to be in a bearish trend if there is a greater available supply than there is demand.

3. ANALYSIS OF STOCK MARKET

The most prevalent application of algorithmic trading is high-frequency trading (HFT). Trading involves buying a potential share at a discount and then selling it at the market's top growth rate. To ascertain the stock's potential, a thorough statistical verification and stock research process is used. The factors that affect this trade activity include time, price, volume, and technical indicators. The trading choice should take human parallax errors into consideration at every stage. In the case of algorithmic trading, these actions are specifically planned to maximize profit while lowering risk in each trading transaction. The algorithm can process in a timely cycle, allowing it to complete more transactions in a given time frame.

3.1 Basic Technical Trading Signals

We can assess market activity and predict future market behaviour thanks to several crucial technological tools of the trade.

- Moving Average Convergence Divergence (MACD): After the trading circumstances have been created, moving average convergence and divergence (MACD) indicators give indications. known also as a trailing signal.

- Aroon Indicator: The Aroon Indicator tracks the new highs and lows in the price movement of the market trend.

- The Average Directional Index (ADI): A price moving trend's momentum and intensity are gauged by the Average Directional Index (ADI). The directional strength is strong when the ADX value is more than 40. When it drops below 20, the vigour is waning.

- Accumulation or Distribution Line (A/D): This measures the volume of trades in a security over time. accumulation and dispersion across either a short, medium, or long distance.

- On Balance Volume (OBV): is a metric that measures the volume of securities across time while accounting for positive and negative flows.

- Relative Strength Index (RSI): These leading signs, known as the Relative Strength Index (RSI), appear before trade-related situations.

- DMI (Directional Movement Indicator): A price indication that contrasts the current share price with the prior price range is known as a DMI (Directional Movement indication). The positive figure denotes an increase in price, whereas the negative value denotes a decrease in price.

- Movement of Trendlines: Over time, a trendline shows the uphill and downhill changes in market values. Create a collection of practises for using trading logistics to purchase and sell stocks on the market. Next, ask your stock broker for API connectivity so you may submit bids right away. Although it might seem simple, it takes a lot of time and effort to create a trading bot that can make millions through high-frequency trading, something that humans cannot do.

3.2 Stock Market Analysis Using Data Science

Everywhere you look, you can read about the power of data science. The problem of data affects everyone. Businesses are interested in learning how data could potentially save expenses and improve their bottom line. The healthcare industry is curious about how data science might help them identify ailments earlier and treat patients more effectively. Data science is frequently represented by numbers. These numbers, however, could apply to anything, such as the number of customers who purchase a product or the quantity of commodities sold. Of course, these numbers could also represent money. In order to develop a distinctive perspective on the stock market and financial data, data science is being used in this manner. Stocks, commodities, and securities all abide by the same basic principles. Over the past 20 years, trading platforms have become more and more popular, but each one has its own features, resources, and fees. Canadians still lack access to zero-commission trading platforms despite this growing trend. In a 12-month study, Gary Stevens of Hosting Canada examined the features that each of the most well-known stock trading platforms provides to its users. To make the greatest decision for you, you must comprehend how they work, and Gary's in-depth explanation may help you with that. We wholeheartedly endorse this article from The Balance if you're interested in learning more about Canadian ETFs.

Recognising Basic Data Science Ideas in the Stock Market In data science, there are a number of idioms that only scientists would comprehend. Data science is just maths plus a dash of programming and statistics skills.

Algorithm: Data science extensively makes use of algorithms. A task-completion algorithm is essentially a series of instructions that must be followed. Probability is that you are aware of the employment of algorithms in the buying and selling of stocks. Using algorithms to decide when to buy or sell stocks is known as algorithmic trading. For instance, a stock may be programmed into an algorithm to be bought if it drops by 8% during the course of the day or sold if it loses 10% of its value since being purchased. Algorithms are designed to function without the assistance of humans. They are commonly referred to as bots. They behave much like robots while making logical decisions.

Training: We're not discussing running a 50-meter dash. Training is the process of using data to educate a system how to react in machine learning and data science. We are able to create a learning model. A computer can now make precise predictions based on historical data thanks to machine learning. You will need a model of the stock values from the previous year to serve as a foundation to predict what will happen if you want to teach a computer to estimate future stock prices.

Testing: The training set would be the data from January to October. Then, we'll run tests in November and December. The predictions made by the computer will be contrasted with actual prices.

The modeling used to Predict Stock Prices: The foundation of data science is modelling. With this approach, past actions are analysed mathematically in order to predict future outcomes. In the stock market, a time series model is used. A time series is a collection of data that is indexed across time, in this case the value of a stock. It is possible to divide this time span into hours, days, months, or even minutes. By gathering pricing data using machine learning and/or deep learning methods, a time series model is created. Before fitting the model to the data, analysis of the

data is required. This is what makes it possible to anticipate future stock values over a certain period of time. A second type of modelling used in machine learning and data science is a categorization model. These models attempt to categorise or anticipate what is represented by the data points when they are provided input. A machine learning model may be given financial information like the P/E ratio, total debt, volume, and other factors to determine whether or not a stock is a good investment when talking about the stock market or stocks in general. Based on the financial data we supply, a model can determine if it is the right time to sell, hold, or buy a stock. It's possible for a model to foresee something so complex that it ignores the relationship between the feature and the desired outcome. Overfitting is the term for this. When a model does not accurately reflect the data, it is said to be underfit, which leads to overly straightforward forecasts. If the model has trouble detecting stock market patterns, overfitting is an issue.When a model predicts the fundamental average price based on the entire history of the asset, underfitting occurs. Poor forecasts and projections are the outcome of both overfitting and underfitting. We have only begun to explore the connections between stock market investment and the ideas of machine learning. But it's important to understand the fundamental ideas we covered today since they lay the groundwork for realising how machine learning is used to predict what the stock market can do. There are more subjects available for those who wish to go into the details of data science and how it relates to the stock market.

4. DATA SCIENCE IN STOCK MARKET

Numerous things can be represented by the numbers that are usually used to represent data. These numbers could be for sales, inventory, clients, and last but not least, money. This brings up financial information, specifically the stock market. Stocks, commodities, securities, and other trading instruments are largely comparable. Data Science uses a lot of terms, expressions, and jargon that many people are not familiar with. We are available to assist with all of that. Data science requires a working knowledge of mathematics, statistics, and programming. I'll provide links to numerous resources throughout the text if you want to read more about these ideas. Let's go right to what we all wanted to know now: how to build market evaluations using data science. To decide whether a stock is worthwhile for investment, we employ analytics. Now let's go over some data science concepts that have to do with money and the stock market. The fields of data science and programming frequently use algorithms. A set of guidelines that must be followed in order to finish a specific activity is known as an algorithm.

A growing trend in the stock market is algorithmic trading, as you may have heard. Trading algorithms used in algorithmic trading include parameters like buying a stock only when it has declined by exactly 5% that day or selling a stock once it has lost 10% of its value since being bought. All of these algorithms have the ability to function without the assistance of a person. Since their trading strategies are largely mechanical and they trade without emotion, they are occasionally referred to as trading bots. Programming, mathematics, and business are all combined in the multidisciplinary field of data science and analytics. You need to understand both words before you can distinguish between them. Therefore, let's start with data science.The term "data science" is used to refer to a wide range of approaches and strategies for obtaining information. In plainer language. Data science is a field that consists of a variety of tools, machine learning strategies, and algorithms for finding patterns in unstructured data. Data analytics is a process for increasing output and revenue. In this part, data sets are examined in order to draw conclusions about the information they contain. Information is extracted, categorised, and subject to various approaches depending on organisational needs in order to discover and assess conduct information. It was also known as data analysis.

Data Scientist: Computer techniques like neural networks and machine learning, as well as knowledge of Applied Statistics, Data Mining, are all necessary. It's vital to be familiar with database architectures like MySQL, Hive, and others. Data science is used in a variety of scenarios, such as online searches and digital advertising. The progress of machine learning and AI depends on data science. Then data analysts construct an algorithm using the data.

Data analysts: The ability to retrieve and query data is necessary. Data blending, data purification, data discovering, and data visualisation are among a data analyst's primary responsibilities. It is crucial to have a basic understanding of statistics. The best industries for data extraction by analysts include healthcare, gaming, and retail.

5. PATTERNS IN STOCK MARKET

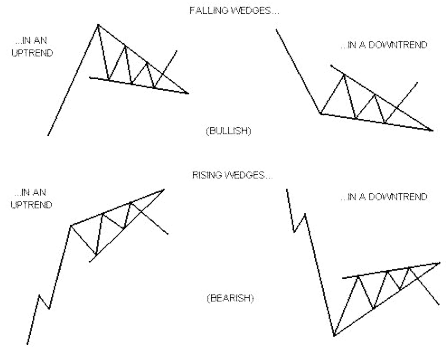
Continuation Patterns: An ongoing trend can be considered to come to an end with a continuation pattern. This happens whenever a bull market pauses during an upswing or a bear market pauses during a downturn. When a pricing pattern begins to take shape, it is impossible to tell whether the trend will last or turn around. To determine if the price breaks above or below the continuation zone, as well as the trendlines that were employed to create the price pattern, pay great attention. Technical analysts frequently predict that a trend will last as long as it cannot be demonstrated to have changed.

Reversal patterns: are pricing patterns that show a change in the direction of the present trend. When the bulls or bears have peaked, these patterns show it. The existing trend will halt before continuing in a new direction as fresh energy emerges from the opposing side (bull or bear).

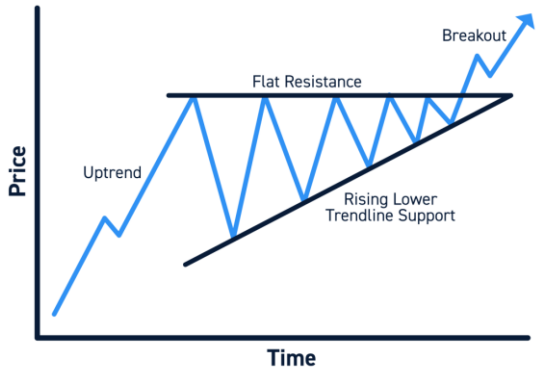
Pennant: Pennants are continuation patterns that are created when two trendlines cross. Pennants are characterized by trendlines that move in opposite directions, one up and one down. Below is an illustration of a pennant. Frequently, the volume will decrease while the pennant is built and then increase when the price eventually breaks through. A bearish pennant pattern denotes a price trend in the negative direction. Volume is decreasing, and a flagpole forms on the right side of the pennant, forming a bearish pattern.

Flag: Flag patterns consist of two parallel trendlines that may be inclined upward, downward, or horizontally. In contrast, a flag with a downward bias (bearish) shows as a break in an uptrending market. A flag with an upward slope (bullish) appears as a standstill in a downtrending market. Usually, when a flag form develops, there is a reduction in volume that is followed by a rise in volume when the price breaks out of the flag shape.

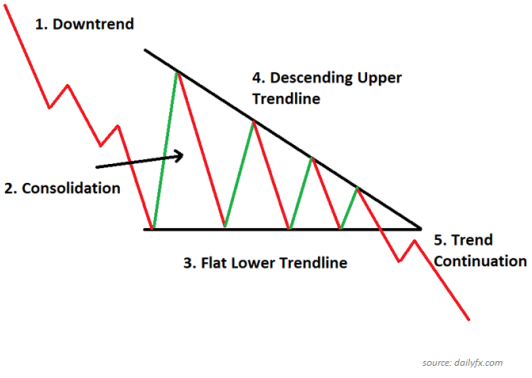
Wedge: Pennants and wedges are both continuation patterns made of two trendlines that are convergent; however, a wedge differs from a pennant in that both trendlines are moving in the same direction, either up or down. A wedge with a downward slope denotes a break in an upswing, whereas one with an upward slope denotes a pause in a downturn. while with pennants and flags, volume frequently declines while patterns form before rising after the price breaks above or below the wedge pattern.



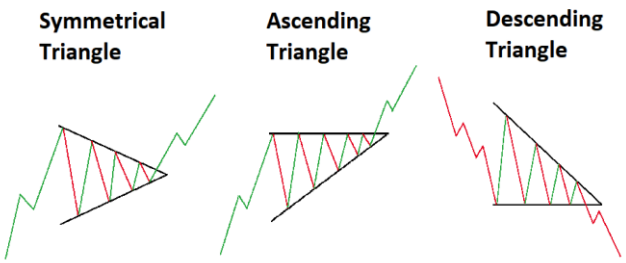
Ascending Triangle: A trend continuation pattern with a stated entry point, profit target, and stop loss level is an ascending triangle. The entrance point is where the resistance line crosses the breakout line. An ascending triangle is a bullish trading pattern.



Descending Triangle: A descending upper trend line suggests that a collapse is about to occur, and a descending triangle is the opposite of an ascending triangle in that it suggests declining demand.



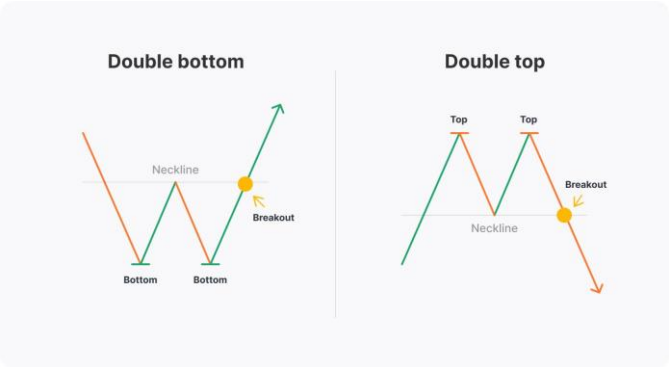
Symmetrical Triangles: There is neither an upward nor downward trend in symmetrical triangles, which emerge when two trend lines converge in their direction and signal an impending breakout. As shown in the illustration below, the size of the breakouts or breakdowns is frequently the same as the height of the triangle's left vertical side.



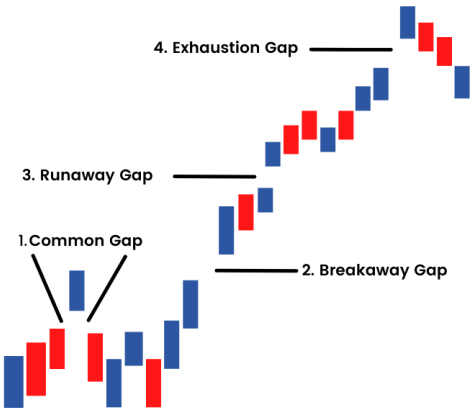
Handle and Cup: A bullish continuation pattern called the cup and handle shows that an upward trend has halted but will pick up again if the pattern is confirmed. The "cup" portion of the design should be a "U" shape that resembles the rounding of a bowl rather than a "V" shape with equal highs on both sides of the cup.

Head and Shoulders: A reversal pattern known as the head and shoulders can appear at market peaks or bottoms as a series of three pushes: an initial peak or trough, a second, larger one, and a third push that repeats the first.A head and shoulders top pattern has the potential to reverse an uptrend and start a decline. It is almost certain that an upward trend will resume after a decline that forms a head and shoulders bottom (or an inverted head and shoulders). As seen in the image below, trendlines can be built to connect the peaks and valleys between the head and shoulders. These trendlines can be horizontal or slightly slanted. Volume may decrease while the pattern takes shape and then rise after the pricing pattern is broken.

Double Top and Bottom: These reversal patterns show areas where the market has twice failed to break through a support or resistance level. They are used to identify these areas.A double top is characterised by an initial push-up to a resistance level, followed by a second unsuccessful attempt, which results in a trend reversal. It commonly resembles the letter M.



Gaps: Patterns of reversal are gaps. They appear if a significant price increase or decrease occurs between two trading periods. For instance, a stock may close at $5 and open at $7 after great results or other news. The three primary categories of gaps are fatigue, runaway, and breakaway gaps. Breakaway gaps, runaway gaps, and fatigue gaps all appear during the beginning, middle, and end of trends, respectively.



6. CONCLUSION

Technical analysis seeks to forecast future changes in financial price based on past price movements. Because it doesn't generate precise predictions about the future, think of technical analysis as being analogous to weather forecasting. Technical analysis, on the other hand, can help investors make predictions about what will "likely" happen to prices over time. Stocks, indexes, commodities, futures, or any other tradable asset whose price is influenced by supply and demand dynamics can all be subject to technical analysis. Price data is any combination of the open, high, low, close, volume, or open interest for a certain asset over a

specific period of time (or, as John Murphy prefers, "market action"). The price information might be daily, weekly, or monthly as well as intraday (1-minute, 5-minute, 10-minute, 15-minute, 30-minute, or hourly). Many techniques such as Machine Learning and Data Science can be used to predict Stock Market but it can and will never be predicted with 100% accuracy.

REFERENCES

[1] S. Kamil, A. Pinar, D. Gunter, M. Lijewski, L. Oliker, J. Shalf, "Reconfigurable Hybrid Interconnection for Static and Dynamic Scientific Applications", ACM International Computing Frontiers, 2007.

[2] IPM Homepage: http://ipm-hpc.sourceforge.net/

[3] C. Bell, D. Bonachea, R. Nishtala, K. Yelick, "Optimizing Bandwidth Limited Problems Using

[4] Robert W. Wisniewski, Peter F. Sweeney, Kartik Sudeep, Matthias Hauswirth, Evelyn Duesterwald, Calin Cascavel, and Reza Azimi, "Performance and Environment Monitoring for Whole-System Characterization and Optimization", PAC2 (Conference on Power/Performance interaction with Architecture, Circuits, and Compilers), 2004.

[5] G. Bell, J. Gray, A. Szalay, "Petascale Computational Systems," IEEE Computer, Volume: 39, Issue: 1 Jan. 2006, pp 110-112.

[6] William T.C. Kramer, John Shalf, and Erich Strohmaier, "The NERSC Sustained System Performance (SSP) Metric," LBNL Technical Report LBNL-58868, September 2005.

[7] S. Kamil, J. Shalf, L. Oliker, and D. Skinner, "Understanding Ultra-Scale Application Communication Requirements," in Proceedings of the 2005 IEEE International Symposium on Workload Characterization (IISWC), Austin, TX, Oct. 6–8, 2005, pp. 178–187. (LBNL-58059)

[8] Krste Asanovic, et. al., "The Landscape of Parallel Computing Research: A View from Berkeley, Electrical Engineering and Computer Sciences," the University of California at Berkeley, Technical Report No. UCB/EECS-2006-183, December 18, 2006. (http://view.eecs.berkeley.edu/)