2013

Forecasting Microprocessor Technology in the Multicore Era Using TFDEA

Saranya Durairajan  
*Portland State University*

Maria Ibarra Prado  
*Portland State University*

Noshad Rahimi  
*Portland State University*

Shabnam Razeghian Jahromi  
*Portland State University*

Let us know how access to this document benefits you.

Follow this and additional works at: [http://pdxscholar.library.pdx.edu/etm_fac](http://pdxscholar.library.pdx.edu/etm_fac)

Part of the Engineering Commons

Citation Details

Durairajan, Saranya; Prado, Maria Ibarra; Rahimi, Noshad; and Jahromi, Shabnam Razeghian, Forecasting Microprocessor Technology in the Multicore Era Using TFDEA, Portland International Conference on Management of Engineering and Technology (PICMET), Portland, OR, 2013.

This Article is brought to you for free and open access. It has been accepted for inclusion in Engineering and Technology Management Faculty Publications and Presentations by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.
Abstract—Technological advancements in the microprocessor industry are benchmarked and gauged against a set of diverse criteria, specific to the fabrication process, usage as well as achieved performance. Changing trends in the appeal factor as well as wide variety of growing application of microprocessors in different industries also have a defining impact in the advancement of the technological features in future. This study improves the previous investigation in forecasting microprocessors’ technology and uses Technology Forecasting using Data Envelopment Analysis (TFDEA) methodology for an enhanced model. The study takes advantage of the recent microprocessor dataset including multi core processors, from a variety of resources including the dataset collected by Stanford University, the database made available by Standard Performance Evaluation Corporation (SPEC), and the specifications announcements by the microprocessor manufacturers such as Intel and AMD. The result of this study is a rate of change (RoC) that is obtained based on the recent design trends including the State-of-the-Art generation of multi-core microprocessors and hence, is superior for forecasting the future microprocessor technology trends. The Rate of Change obtained provides the rate in which values of expected output performance or input requirements for the state of the art microprocessors change in future years and can be used to evaluate the competitiveness of the projects being researched and developed.

I. INTRODUCTION

Moore's law has been the yardstick for technology advancement of microprocessors, predicting that the number of transistors on integrated circuits doubles approximately every two years[1]. Such trajectory has proven to be persistent over the past several decades and has provided a reliable technology roadmap for semiconductor industry; doubling the number of transistors every two years to achieve the expected performance. The number of transistors, however, is only one aspect in complexity of designing a high performance microprocessor. In the recent market, designers are challenged to come up with innovative ways to design high speed microprocessors that best fit different applications including high performance with minimum power consumption. In such application one can forego high output performance for a more prudent input. Anderson et al. suggested an alternative to Moore’s law where they incorporated a wider features of microprocessor design complexity and used Data Envelopment Analysis (DEA) as a means to measure technological progress over time[2]. Transistor count, Power, Minimum feature size, Die size and SPEC CPU benchmarks were used as parameters to measure the microprocessors’ State of the Art (SOA) using data from 1990 – 1999 time periods.

The earlier study was limited to single core processors and was using an older dataset. The results obtained by Anderson et al showed a slower rate of technological progress than would be expected from Moore’s law, mainly due to the difficulties imposed by the feature size and die size reduction.

The objective of this paper is to extend the dataset to include processors between 1998 and 2012 and use Technology Forecasting using Data Envelopment Analysis (TFDEA) methodology to forecast and estimate the availability of future characteristics of microprocessors. TFDEA makes use of the DEA technique, which provides the capability to analyze multiple inputs and outputs simultaneously and produces an efficiency frontier that contains the best performers[3].

II. LITERATURE REVIEW

A. Technology Trends

In the past 20 years, microprocessors technology has experienced improvements in circuit integration and microprocessor throughput. The technology has grown rapidly due to transistor speed, energy scaling and core micro architecture advances powered by Moore’s law. In every generation (two years), transistor density has doubled as their dimensions have been reduced by 30% (shrinking their area 50%), and circuits have become 40% faster increasing the whole system performance[4]. However, due to battery capacity and chip reliability (heat dissipation limits), power consumption has been the key limiting factor for performance scaling in the single-core microprocessor technology.

In the past decade, multi-core microprocessors have become the major design trend. Limits in instruction level parallelism (ILP) and power dissipation constraints have triggered the high performance microprocessor roadmap to enter the multi-core era, starting from the high-end server processors and moving to the low-end hand-held mobile device processors. A multi-core micro architecture provided an effective alternative to improve throughput performance of parallel programs while keeping power consumption under the control. To improve efficiency, single-thread performance was sacrificed and instead multiple cores were joined on a single chip when more transistors became available. The more threads accommodated in the application set, the more efficient the processors became [5], [6]. Recently the typical pattern among multi-core CPU products is to keep the number of cores constant within a generation and double the number of transistors within each core [7]. By exploiting Moore’s Law to replicate cores, multi-core architectures increased computational performance. However, there is no real benefit if the software has no parallelism [8].
Core micro architecture techniques took advantage of the abundance of transistor integrity to deliver improved performance; nevertheless growing power densities are still the major constraint to performance improvements. Initially, multi-core processors were designed with a step back in core-level frequency allowing throughput increase at affordable power; however the power consumption and dissipation problem did not disappear with the multi-core era [9]. With a flat power budget, from mobile platforms to PCs and workstations to the largest supercomputers being all power limited, power efficiency is one of the primary metrics for, and driver of, microprocessor designs[6], [9]. Power and heat management are the two major concerns that are more pronounced with the addition of multiple cores.

As power continues to limit performance scaling, researchers forecast that processor designs will be forced to use large-scale parallelism and heterogeneous cores (application-customized), or a few large cores and a large number of small cores operating at low frequency and low voltage, as alternatives to achieve performance and energy efficiency [6], [9].

B. Forecasting Tools
Technology forecasting is the act of forecasting inventions, innovations, or diffusion of technologies. It is a procedure of collecting data and analyzing them to predict future technological developments and its social effects [10], [11]. It is a popular technique among companies because it can be used to design future products to outperform competitors. Specifically, technology forecasting provides companies with a capability of studying the impact of past inventions, innovations, or diffusion of technologies. It is a procedure of collecting data and analyzing them to predict future technological developments and its social effects [12], [13], [14], [15], [16]. Limitations include,

- Single variable based prediction: Technology is impacted by different attributes; it is hard to find the sole characteristic/variable that will impact technology in future
- Preference changes over time is not considered, thereby, unsuited for dynamic trade-offs
- Correlation between technology attributes is not considered. Technology attributes are assumed independent; for less known technologies it is difficult to detach the attributes.
- Lack of a multiple output model. Current methods work with a single output at a time; the outputs are fixed and there is no ability to waive any of them [12].

Technology forecasting using Data Envelopment Analysis (TFDEA) is recognized as a powerful forecasting method in literature that addresses the above gaps. It is a non-parametric method that can incorporate multiple inputs and outputs to identify the best performers at each observation period and forecast the technology trend accordingly. It does not require a mathematical specification of functional relations between inputs and outputs [12]. Technology forecasting via DEA is however very sensitive to the choice of variables. Therefore the inputs and outputs parameters need to be carefully selected.

III. RESEARCH METHODOLOGY
TFDEA is the forecast method used in this research. It evaluates the microprocessor technology’s historical stages against the State-of-the-art so that the characteristics of the technology’s future can be identified. The first step is to identify the proper decision variables for inputs and outputs. Decision variables are split into structural and functional components. Input variables should indicate the manufacturing, design and usage difficulties; output variables should reflect the performance of a processor [2].

In microprocessors, the parameters - minimum feature size and die area represent the manufacturing difficulty; the number of transistors reflects the design difficulty and power consumption represent the usage difficulty. As the manufacturing process gets more difficult for smaller feature size, the reciprocal of the feature size is used as the input parameter. Due to exponential increase in transistor count over time, \(10^6\) value is used in the model [2]. Thermal Design Power (TDP) is the maximum power that can be dissipated by a microprocessor [17] and is used to reflect the value of the maximum power consumption of the microprocessor.

This study uses two speed metrics for performance. SPEC CPU subcommittee benchmarks the processor performance through two program suites designed to evaluate integer and floating point arithmetic calculations, described as SPECint and SPECfp respectively, in SPEC tests [18]. Speed metrics (SPECint and SPECfp) represent the single-core single-task scenario. Table 1 lists the parameters used in this study to forecast microprocessor technology.

<table>
<thead>
<tr>
<th>TABLE 1: MODEL PARAMETERS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input Parameters</strong></td>
</tr>
<tr>
<td>Die size</td>
</tr>
<tr>
<td>Feature size</td>
</tr>
<tr>
<td>Power (Thermal Design Power)</td>
</tr>
<tr>
<td>Number of transistors</td>
</tr>
</tbody>
</table>
A total of 193 microprocessors, released between 1998 and 2012, from manufacturers such as AMD, DEC, Fujitsu, HP, IBM, Intel and SUN, are included in this study [19]. For parameters listed in Table 1, Stanford CPU database is used as the primary data source. In addition, data from the SPEC and manufacturer websites are used to include more recent microprocessors. Generally, microprocessor performance is measured using the SPEC benchmark suite at the time of their release. For example, Intel Nehalem Clarkdale’s performance is given in SPEC CPU2006 (SPECint2006, SPECfp2006) scores since it was released in 2010; whereas, Dec Alpha 21364, released in 2002, was measured in SPEC CPU2000 benchmark. This study employs the SPEC CPU2000 and CPU2006 performance scores, while converting the 2000 scores to 2006 equivalents (using equations found by regression analysis among the microprocessors that had values for both benchmarks). In both SPECint and SPECfp cases, Equations (1) and (2) were obtained with $R^2$ value of approximately 91%.

$$fp\text{2006 equivalent score } = 0.0053(SPECfp\text{2000}) + 0.804 \quad (1)$$

$$int\text{2006 equivalent score } = 0.0058(SPECint\text{2000}) + 1.054 \quad (2)$$

While these equations help to create a data set with two measures that relate to the output performance, they could not be considered as the true output values needed for the model. The CPU2006 speed scores were performance measures based on single thread scenario, as opposed to test cases that could be parallelized and use the multiple cores in the newer microprocessors. This means that these rates under-evaluated the performance of the multi-core processors in recent years, and hence were not the appropriate scales for this study. In the absence of universal measures in which all the microprocessors across different generations could be compared, this study attempts to calculate the best estimate using Gustafson’s Law and the result is normalized performance metrics among all the processors in the dataset.

### A. Performance Measurement for Multi-core Processors

Gustafson’s Law [20] is used to calculate the proportional speed up measures based on SPECint and SPECfp values. In general, speed up is a function of the number of processors. In an ideal setup, when running an algorithm with linear speedup, doubling the number of processors, doubles the speed.

According to Gustafson's Law, computations involving large data sets can be easily divided into a set of parallel instructions. This Law offers a realistic look at the potential of parallel computing on multi-core processors [20], [21] and is described by Equation (2).

$$S(p) = p - \alpha \cdot (p - 1). \quad (2)$$

Here $p$ is the number of cores; $S$ is the speedup and $\alpha$ is the non-parallelizable fraction index of a parallel process. In this study $\alpha$ is assumed to be 10%. Such assumption is based on the progresses in parallel computing which allows the majority of a program to be executed in parallel [22].

In this research, the above formula is used to calculate the overall performance of a microprocessor. Integer and floating point speedup equivalents calculated as shown in Equations (3) and (4) are used as outputs in this study.

$$int\ speedup = SPECint \cdot [p - \alpha \cdot (p - 1)]. \quad (3)$$

$$fp\ speedup = SPECfp \cdot [p - \alpha \cdot (p - 1)]. \quad (4)$$

### V. TFDEA MODEL IMPLEMENTATION

The proposed TFDEA model is implemented using the tool developed by Lim and Anderson [23], for the inputs and outputs listed in Table 2.

#### A. Model Orientation

In order to determine the efficient (i.e. ‘best practice’) frontier using DEA, one can choose between DEA input-oriented and output-oriented models, based on the objective of the technology under study. An input-oriented model is used when the target for the product under analysis is to minimize its input for a given output. An output-oriented model is used when the scope is output maximization for a given input. Until early 2000, CPUs were able to keep up and even exceed the expectation of doubled performance every 18-20 months. From mid-2000, multi-core computing has become mainstream and single-threaded CPU performance did not scale as before [24], with the key limitation being the power [25]. With increased performance via deep pipelines and superscalar computation, typical high-end microprocessor power went from less than a watt to over 100 watts. Though decreased feature size aided reduced power, with the addition of large amount of logic in modern-day microprocessors, coupled with increased operation frequency, the overall effect was an exponential increase in power by each subsequent processor generation. In the current market, since the primary focus is on reducing power, an input oriented model is used in this study.

#### B. Frontier Year

Figure 1 shows the number of microprocessors as of each study year in the dataset. As it shows till 2006, relatively there were not enough DMUs available for the forecast, and hence, the frontier year had to be chosen from year 2006 onwards.

<table>
<thead>
<tr>
<th>TABLE 2: Inputs &amp; Outputs of the TFDEA Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inputs</td>
</tr>
<tr>
<td>Power (TDP) [Watts]</td>
</tr>
<tr>
<td>Die size [mm²]</td>
</tr>
<tr>
<td>Reciprocal of Feature size [µm⁻¹]</td>
</tr>
<tr>
<td>Logarithm 10 of Transistor count</td>
</tr>
<tr>
<td>Outputs</td>
</tr>
<tr>
<td>Int speedup</td>
</tr>
<tr>
<td>Fp speedup</td>
</tr>
</tbody>
</table>

2110
A proper frontier year is the one that is recent enough so that the rate of change calculated by TFDEA includes the recent technological advances. Also, there must be enough number of DMUs after the year of frontier, to verify the validity of the model’s forecast. To back test the model, frontier years from 2008 to 2012 are used in this study.

VI. RESULTS

The model is evaluated via back testing against the historical data. The historical data comprises of two sets for different frontier years (Table 3). After detailed analysis, following model is chosen based on superior results.

- Input Orientation
- Constant Returns to Scale
- Frontier years: 2008 to 2012

Model results are compared against the actual historical data for this period. Both the RoC (Rate of Change) and Median Absolute Deviation (MAD) values are calculated as shown in Table 3.

Table: ROC & MAD Values

<table>
<thead>
<tr>
<th>Frontier Year</th>
<th>Learning Period</th>
<th>Validation Period</th>
<th>RoC</th>
<th>MAD [years]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>1998 – 2012</td>
<td>2012</td>
<td>1.244552</td>
<td>0.700999</td>
</tr>
</tbody>
</table>

Table: Parameter Values of the Outlier

<table>
<thead>
<tr>
<th>Name</th>
<th>Release Date</th>
<th>Cores</th>
<th>Power</th>
<th>Log 10 of Trans</th>
<th>Die Size</th>
<th>Rec of Feature Size</th>
<th>Int Speedup</th>
<th>FP Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>IntelCore i5Nehalem</td>
<td>2010</td>
<td>2</td>
<td>35</td>
<td>8.58</td>
<td>81</td>
<td>31.25</td>
<td>38.00</td>
<td>43.13</td>
</tr>
<tr>
<td>IntelCore i7Nehalem-610E</td>
<td>2010</td>
<td>2</td>
<td>35</td>
<td>8.58</td>
<td>81</td>
<td>31.25</td>
<td>41.04</td>
<td>45.79</td>
</tr>
<tr>
<td>IntelPentiumNehalem</td>
<td>2010</td>
<td>2</td>
<td>73</td>
<td>8.58</td>
<td>81</td>
<td>31.25</td>
<td>42.51</td>
<td>48.47</td>
</tr>
<tr>
<td>IntelCore i7Nehalem-980x</td>
<td>2010</td>
<td>6</td>
<td>130</td>
<td>9.07</td>
<td>248</td>
<td>31.25</td>
<td>193.33</td>
<td>207.08</td>
</tr>
<tr>
<td>IntelXeonNehalem</td>
<td>2010</td>
<td>2</td>
<td>30</td>
<td>8.58</td>
<td>81</td>
<td>31.25</td>
<td>41.23</td>
<td>45.66</td>
</tr>
</tbody>
</table>

Figure 2 depicts the MAD and RoC values obtained for the different frontier years. Figures 3 to 7 represent the forecasted results obtained by comparing the forecasted date to the actual release date for all the microprocessors used in the model for frontier years 2009–2012 respectively. The red line in the graphs represents the ideal forecast, i.e. when the model prediction matches the actual dates of release. The area above the red line shows the products that were produced earlier than the model predicted, and the area below the red line shows the ones produced after the forecast date.

For the frontier year 2008, a calculated RoC of 46% is due to the fact that in 2006 and 2007, the release of many multi-core processors caused a dramatic shift in the inputs and outputs compared to previous years. Since this is a very aggressive RoC, most of the microprocessors forecasted using this frontier year, are observed to be released after the forecast date. For years 2009 to 2012, this dramatic shift in the RoC is absent.

For frontier year 2011, the 6 core Intel i7-980x from Nehalem family was removed from the dataset. The i7-980x was very unique at the time of release; it had 9 billion transistors that is 10 times more compared to the transistors’ count of other microprocessors from the same year and much higher power consumption of 130 watts with an overwhelming performance. The microprocessor seemed to be an outlier in 2010 and a big scale one for a different market. Keeping this microprocessor would cause the SOA of 2010 to go obsolete within 14 days and such radical advancement would result in a very high RoC that is not reasonable. Table 4 shows the input and output parameters of 980x and its peer microprocessors. As the table depicts, the values of the parameters for the 980x is significantly higher than the other microprocessors in the same year.
Another observation that stands out in Figure 2 is the significantly higher MAD (1.47 years) when using frontier year 2011. This MAD could be improved and be more aligned with former MAD values by keeping the 980x and removing the 610E instead. In this case, the MAD would be 0.61, which is much closer to the RoC of 2012, and 2010 years (0.7 and 0.86 respectively). Keeping the 980x will result in an increase in RoC (from 1.21 to 1.26), which will improve the accuracy in forecasting SOAs of 2012. However, the i7 610E is not an outlier and removing it did not make sense. Therefore, in this study the RoC 1.21 and the MAD 1.47 have been considered to back test the model. For frontier year 2012 and calculation of the final RoC, the Intel i7-980x was included.

From Table 3, it is found that the model using frontier year 2009 gave the least MAD. But, since 2012 is the latest year of the study, the RoC obtained using the frontier year 2012 is selected for future predictions.
Figure 1: Forecasted Date vs. Release Date for frontier year 2010

Figure 6: Forecasted Date vs. Release Date for frontier year 2011

Figure 7: Forecasted Date vs. Release Date for frontier year 2012
VII. IMPLICATIONS

TFDEA model for the microprocessor technology gave accurate forecast with MAD of less than a year for most of the cases. Based on the above results and the RoC of frontier year 2012, the future SOA products can be extrapolated. Since the model uses an input-oriented measure, input characteristics can be multiplied by the average RoC, assuming constant output characteristics. Hence, in case of maintaining same performance levels, future microprocessors are expected to reduce their maximum power consumption by 24.455 ± 2.49 percent (i.e., 21.965% to 26.945%), using a confidence interval of 95%.

**TABLE 4: ROC AND SD OF FRONTIER YEAR 2012**

<table>
<thead>
<tr>
<th>Rate of Change</th>
<th>1.244552</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviation</td>
<td>0.09375342</td>
</tr>
<tr>
<td>95% Confidence Interval</td>
<td>± 0.02490482</td>
</tr>
</tbody>
</table>

Table 6 shows the results of TFDEA software using 2012 frontier. All the microprocessors shown in this table are from the Ivy Bridge family. Although the performance increase is about 10% compared to the Sandy Bridge family in 2011, Ivy Bridge is using the 22nm design process, which is efficient compared to former 32nm architecture. Intel Core i7-3770T and i7-3770S being the frontiers of 2012 are members of low power series with TDP values of 45 and 65 respectively. Intel Core i5-3570T is also the member of low power models of Ivy Bridge i5 series [26].

Table 7 shows the current input values of Intel Core i7-3770T. The parameter values of i7-3770T and the RoC in Table 4 can be used to project the characteristics of the frontier microprocessors in the future. According to our research, the frontier microprocessors input parameters should decrease by 24% assuming the output parameters remain constant. This projection will be an input-oriented one focusing on decreasing input parameters like power which is the current focus of semi-conductor industry. An output-oriented projection can also be done by keeping the inputs constant and increase the output parameters by 24%. In the case of microprocessors a radial projection is not practically feasible. Moreover, feature size of the microprocessors is following the lithography roadmap and the future values are already determined. Table 7 is showing the future characteristics of a frontier microprocessor using a non-radial projection and the forecasted feature-size in 2013 and 2014 based on ITRS reports [7].

By 2014, an SOA microprocessor is expected to have power consumption of 25.2 watts, die size of 130 mm² and feature size of 18 nm assuming a performance value identical to that of Intel i7-3770T. Any increase in any one of these values, must be compensated by a decrease in other parameters or an increase in the performance. As discussed the projection is a non-radial one as decreasing number of transistors is not likely to happen, especially when the speed-up parameters are assumed to remain constant. These kinds of forecasts provide a very valuable scale for the decision makers in the Microprocessor manufacturing industry. During the early stage of evaluation, using the Rate of Change value, a manager can readily evaluate the target specification against the expected performance at the time of the release. If the target specification is not at least at the same level as the forecasted value, this indicates that the product will not be as good as those released by the competitors at the time of the release and the project should not be pursued.

Additionally, in early stage of the research and development, the Rate of Change can contribute great insights as to, at the minimum level, what the product specification should be when released. If the specification, for any reason, cannot or is not to be improved, then this forecasting method provides a maximum duration of the project to release the product to be marketable and not behind the typical performance of the similar processors in the market.
VIII. CONCLUSION AND FUTURE WORK

This research utilizes the TFDEA methodology to forecast the trend of microprocessors’ technology. The dataset used to calculate the Rate of Change (RoC) consists of both single core and multi core processors from 1998 to 2012. In order to appropriately capture the multi-core scenario, Gustafson’s Law is applied to generate the normalized speed metrics.

In this research, a consistent RoC of less than 50% (24.455 ± 2.49 percent for the 2012 frontier year) is observed, confirming the difficulties in achieving the desired performance and energy efficiency. Since this RoC is calculated based on a wider and a more recent study period, including a multitude of multiprocessors, it can be used to forecast the future microprocessor technology trends in an improved manner.

Due to limitations on the available data, the current study uses TDP values, based on the assumption of energy conservation which states that “the energy expended per instruction as the instruction is processed in the microprocessor pipeline from fetch, decode, schedule, execute, to retirement; is the same amount of energy dissipated as heat” [27]. Future work can include actual power consumption values for the input, instead of TDP values.

Additionally, the current dataset can be expanded further to include variety of modern generation of microprocessors including those designed for handheld devices like ARM technology. A similar study can be performed with a new performance measure (e.g., next generation SPEC benchmark) that can better compare and scale the future generations. The future study can potentially include a universal performance measure (if one becomes available) that will allow to compare performance across different categories of processors including those used in handheld and mobile devices, as well as high performance servers and supercomputers.

REFERENCES