

Article

A Novel Statistical Learning-Based Methodology for Measuring the Goodness of Energy Profiles of Applications Executing on Multicore Computing Platforms

Muhammad Fahad , Arsalan Shahid , Ravi Reddy Manumachu  and Alexey Lastovetsky 

School of Computer Science, University College Dublin, Belfield, Dublin-4, Ireland;

arsalan.shahid@ucdconnect.ie (A.S.); ravi.manumachu@ucd.ie (R.R.M.); alexey.lastovetsky@ucd.ie (A.L.)

* Correspondence: muhammad.fahad@ucdconnect.ie; Tel.: +353-1-716-2521

Received: 29 June 2020; Accepted: 28 July 2020; Published: 1 August 2020



Abstract: Accurate energy profiles are essential to the optimization of parallel applications for energy through workload distribution. Since there are many model-based methods available for efficient construction of energy profiles, we need an approach to measure the goodness of the profiles compared with the ground-truth profile, which is usually built by a time-consuming but reliable method. Correlation coefficient and relative error are two such popular statistical approaches, but they assume that profiles be linear or at least very smooth functions of workload size. This assumption does not hold true in the multicore era. Due to the complex shapes of energy profiles of applications on modern multicore platforms, the statistical methods can often rank inaccurate energy profiles higher than more accurate ones and employing such profiles in the energy optimization loop of an application leads to significant energy losses (up to 54% in our case). In this work, we present the first method specifically designed for goodness measurement of energy profiles. First, it analyses the underlying energy consumption trend of each energy profile and removes the profiles that exhibit a trend different from that of the ground truth. Then, it ranks the remaining energy profiles using the Euclidean distances as a metric. We demonstrate that the proposed method is more accurate than the statistical approaches and can save a significant amount of energy.

Keywords: energy efficient computing; accurate energy modelling; green computing; similarity matching; pattern recognition; anomaly detection

1. Introduction

Energy is identified by the International Energy Agency (IEA) as a major contributor to climate change [1,2]. Energy efficiency is central to the efforts of IEA to combat climate change [3]. Information and communications technology (ICT) systems and devices are predicted in the worst-case scenario to use up to 51% of global electricity in 2030 and contribute up to 23% of globally released greenhouse gas emissions [4]. Therefore, energy efficiency in ICT is becoming a grand technological challenge and is now a first-class design constraint in all computing settings [5,6].

Energy efficiency in ICT can be achieved at the hardware level (or system level) and software level (or application level). While the system-level energy optimization approach focuses on minimizing the energy consumption of the whole node by employing techniques such as clock and power gating, dynamic voltage, and frequency scaling, etc. [7–9], application-level energy optimization techniques use application-level models and model variables such as workload distribution, number of processes, number of threads, etc. [10,11] as decision variables for energy optimization of applications.

Accurate energy profiles as functions of the workload are essential to the optimization of parallel applications for energy through workload distribution [12]. There are many model-based methods for efficient construction of energy profiles but none of them is accurate in all situations. Therefore, to pick the best method in a given situation, we need a way to measure the goodness of energy profiles produced by different methods when the ground-truth profile, often built by a time-consuming and expensive but reliable method, is available. We define the goodness as the accuracy of a profile against the ground truth profile. Here, the ground-truth refers to the baseline profile or the reference value for the comparison. State-of-the-art but inaccurate energy measurements used in energy optimization of applications can result in significant energy losses [13], up to 84% in some real-life settings [14].

Pearson correlation coefficient [15] and average prediction error (also known as relative error) [16] are the most commonly used statistical measurements to determine the accuracy of energy profiles. A plethora of research work including [5,13,14,17–19] use average, maximum and minimum prediction errors to determine the accuracy of energy profiles. References [20–23] are some of the notable works which used the correlation coefficient to determine whether the energy profiles follow the ground truth.

However, there are research works questioning the effectiveness of both techniques for using goodness measurement of energy profiles. For example, Rico-Gallego et al. [24] argue that the relative error is lower for a profile that underestimates than for a profile that overestimates, and thus can negatively impact the interpretation of the results. Similarly, Fahad et al. [13] demonstrate that the two statistical measures do not capture the holistic picture of the energy consumption trend of the profiles, and thus are blind to the qualitative differences of the energy profiles and the ground truth.

In general, both popular statistical techniques are highly sensitive to outliers and rely on the assumption of linear or smooth increase of energy consumption by applications with the increase of workload. However, the energy profiles of applications on modern multicore platforms are highly non-smooth and non-linear. Therefore, the existing statistical measures can rank an inaccurate energy profile higher than accurate ones. The reason is two-fold. First, in the presence of significant variations in the energy profiles, they do not capture the difference in the general trend of energy consumption. Second, they do not capture the similarities in variations.

While the general direction of energy profiles of applications on multicore platforms is reported as a near-linear increasing function of workload size, the shape of the profile can be highly non-linear and non-smooth [11]. We distinguish the terms *trend* and *shape* using the following example. Consider the sample energy profiles shown in Figure 1. The general direction of all three profiles, which represents the underlying energy consumption trend, is increasing with the increase in workload. However, their shapes are different. The energy profile *Model1* is linear whereas the shapes of *Real* and *Model2* are non-linear and non-smooth.

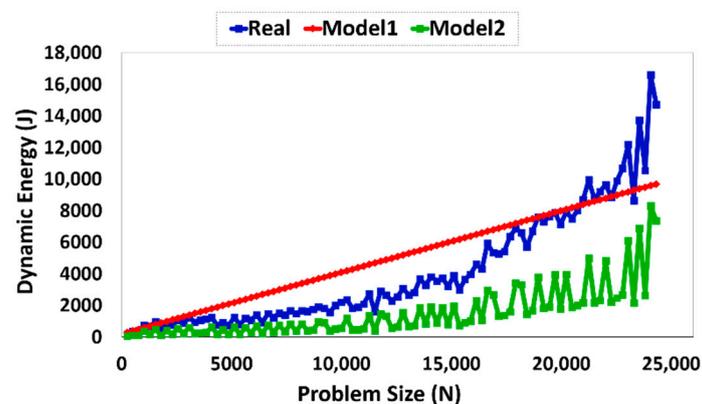


Figure 1. Sample dynamic energy consumption profiles.

While the goodness measuring problem is comparatively less studied for the energy of computing, a plethora of methods have been proposed to solve this problem in many other fields such as data mining, time series similarity analysis, and graph (matching) theory. Popular similarity measures for pattern matching are cosine similarity [25], dynamic time warping [26], angular metric for shape similarity (AMSS) for time series data [27], and autoregressive integrated moving average (ARIMA) method [28–32]. Distance metrics used to determine the pattern matching include Euclidean distance [33–35] and graph-edit-distance (GED) [36]. We provide a brief overview of the popular approaches in these faculties and why they are not applicable straightforwardly for determining the goodness of energy profiles in the Supplementary Materials [37].

To summarize, there is no effective metric to measure the goodness of energy profiles. In this work, we present a novel methodology called trend-based similarity measure (TSM) of energy profiles, which measures the similarity between a given energy profile and the ground truth. TSM is designed to capture the underlying energy consumption trend of the profiles and is composed of the following four stages: (i). The regression model of the energy profile is learned, (ii). The regression fits of this energy profile and the ground truth are compared to determine if they exhibit the same trend, (iii). If they do not, then the energy profile is branded fundamentally inaccurate; (iv). If they do, the distance between the regression models of the energy profile (that follows the same trend-line as of the ground truth) and the ground truth is determined using Euclidean distance as a metric of goodness of the energy profile.

To the best of our knowledge, this is the first work to estimate the goodness of energy profiles by taking into consideration the qualitative difference of the underlying energy consumption trends. Additionally, unlike other statistical methods used for goodness estimation, it uses the Euclidean distance metric for quantitative estimation of similarity between non-linear and non-smooth profiles, increasing the accuracy of estimation. We compare TSM with popular approaches such as Euclidean distance, average and maximum prediction errors, and correlation coefficient for a diverse set of 235 application energy profiles obtained on multicore heterogeneous hybrid computing platforms using three popular energy measurement approaches (i). System-level measurements using power meters, (ii). Integrated on-chip power sensors, and (iii). Energy predictive models. It is shown that the popular statistical approaches do not capture the underlying energy consumption trend and thus erroneously rank an inaccurate energy profile as better than more accurate ones in some cases. We demonstrate that using inaccurate profiles, obtained with state-of-the-art measurement tools, in energy optimization loop may lead to significant energy losses (up to 54% in our case). We find TSM to be more effective when employing in the energy optimization loop than the popular statistical approaches. In summary, the main contributions of this work are:

1. A novel methodology to measure the similarity between an energy profile and the ground truth. To the best of our knowledge, the proposed methodology is the first work that takes into consideration the qualitative differences of the energy consumption trend of the profiles and ranks the energy profiles based on their similarity with the ground truth.
2. An experimental validation of the proposed methodology for a diverse set of 235 application energy profiles on modern multicore hybrid heterogeneous computing platforms.
3. A comprehensive comparative analysis of the proposed methodology with popular statistical approaches such as correlation, average error, and Euclidean distance, which are commonly used to compare the accuracy and similarity of energy profiles as well as time series of equal lengths in general. We demonstrate that all three statistical approaches fail to capture the qualitative difference of an energy profile and the ground truth, and thus fail to distinguish the energy profiles based on their energy consumption trend. Therefore, they can mislead to consider as similar the energy profile whose energy consumption trend is different from that of the ground truth.
4. We demonstrate how the proposed methodology can help in determining whether the energy model that is used to construct the energy profile, includes an extraneous contributor that does not reflect the energy consumption by the application, or it lacks some essential contributor to energy consumption by the application.

5. We compare the effectiveness of our proposed method with state-of-the-art statistical approaches for energy optimization. We demonstrate that the use of the state-of-the-art instead of TSM in the energy optimization loop leads to significant energy losses (up to 54% in our case).

The rest of the paper is organized as follows. We present the proposed solution method in Section 2. The experimental results and a general discussion are presented in Section 3. Finally, we conclude the paper in Section 4.

2. Materials and Methods

This section is organized as follows. We start with the formulation of the goodness measuring problem for energy profiles which are constructed using different energy estimation approaches. We then give an overview of the state-of-the-art goodness of fit techniques followed by a study on their inadequacies to determine the goodness of energy profiles of the applications executing on multicore computing platforms. Then, we describe our proposed solution method TSM. Finally, we explain the experimental platform, the dataset of applications used in this work, and the experiment methodology used to validate TSM.

2.1. Goodness Measuring Problem Formulation

Accuracy or goodness of energy profiles of an application can be defined as the degree to which the energy consumption data of the methods which are employed to produce these profiles, conform to the ground truth. Hence, the similarity of an energy profile with ground truth is also implicitly determined when calculating its accuracy. An energy profile of an application is represented as a function of workload size. We define the *goodness* as a measure that provides an absolute value of resemblance between two vectors (ground truth and an energy profile) in a solution space (i.e., set of energy profiles (EPS) of an application).

Goodness Measuring Problem: Consider an energy profile $E(A)$ of an application A given by a discrete set, $E(A) = \{e(x_1), e(x_2), \dots, e(x_n)\}$ where $e(x_i) \in \{1, 2, \dots, n\}$ is the energy consumption by the workload size x_i . Let there be m energy profiles of the same application A for the same range of problem sizes constructed with different energy measurement approaches. Let EPS_A denotes the set of energy profiles of an application A . Then, the problem is to find the best energy profile in EPS_A which has maximum resemblance with ground truth among all energy profiles.

2.2. Challenges With State-of-the-Art Practices to Measure the Goodness of Energy Models

Multicore architectures are now prevalent in all computing settings ranging from a handheld mobile device to HPC computing platforms and supercomputers. However, the advent of the multicore era has also introduced several inherent complexities, which are severe contention for shared on-chip resources due to the tight integration of tens of cores, non-uniform memory access (NUMA), and dynamic power management (DPM) of multiple power domains such as CPU sockets, Dynamic random-access memory (DRAM). The functional relationships between the energy and workload size have complex (non-linear and non-smooth) properties on modern multicore CPUs. Profile-based energy optimization algorithms [11,12] leverage the profile variations to find energy-efficient workload distributions. At the same time, the state-of-the-art statistical approaches consider energy profiles as linear or smooth functions of workload sizes to find their goodness. The failure of capturing the qualitative differences in energy profiles can drastically affect the energy optimization efforts and can cause significant energy losses [13,14].

We present two case studies to highlight the inadequacies of state-of-the-art statistical approaches. The first is based on the results of [13]. Consider an energy profile of multiplication of two dense $N \times N$ matrices on an Intel Haswell server comprising of two CPU 12-core sockets, using Intel Math Kernel Library routine to compute the Double-precision General Matrix Multiplication (MK-DGEMM) routine (see Figure 2). We run two MKL-DGEMM routines in parallel on both sockets. Each routine solves a

different workload size ($N \times N$). For the first socket, the workload size ranges from $10,000 \times 10,000$ to $14,928 \times 14,928$ with a constant step size of 64. For the second socket, the range is from $15,000 \times 15,000$ to $19,928 \times 19,928$ with a constant step size of 64. The x-axis in the Figure 2 shows the workload sizes for socket1. We measure the total dynamic energy consumption by these two parallel workloads using Intel Running Average Power Limit (RAPL) [38] and HCLWattsUp [39]. The HCLWattsUp Application Programming Interface (API) provides system-level power measurements using external power meters, which is the ground truth.

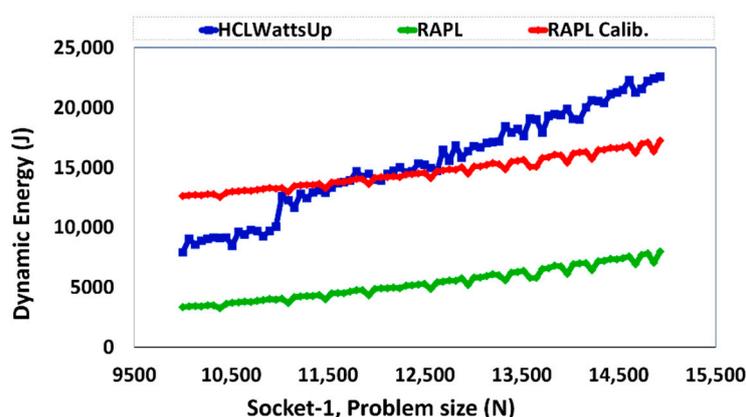


Figure 2. Dynamic energy consumption profile segments of matrix-matrix multiplication using Intel Math Kernel Library (MKL) constructed with HCLWattsUp and Intel Running Average Power Limit (RAPL) on HCLServer01.

RAPL always reports less the dynamic energy consumption, resulting in the average and maximum errors of 64% and 69% respectively. The Euclidean distance of 92,104 between these profiles is large, but there exists a strong positive correlation of 0.97 between them given by the Pearson correlation coefficient. Despite the strong positive correlation, the profiles disagree on energy consumption behavior for almost 50% of the data points. For example, for data points (N) {10512,11152,11984,12624,13712}, HCLWattsUp reports a percentage decrease of {8,5,3,2,6} in dynamic energy consumption with respect to immediately preceding data points. In contrast, RAPL reports a percentage increase of {8,14,9,13,13}. Similarly, while HCLWattsUp suggests a percentage increase for data points {12944,13328,13584,13968} by {3,7,8,3}, RAPL suggests a decrease of {9,7,8,8}.

Furthermore, although both profiles exhibit an overall rising trend of energy consumption, the degrees of the slopes are significantly different. The divergence between the profiles increases with the increase in workload sizes. All three statistical measurements fail to capture this behavior.

The high positive correlation coefficient between the profiles indicates that the RAPL profile can be calibrated. Hence, its average and maximum errors can be reduced to 18% and 59% from 64% and 69% respectively by calibrating the RAPL readings with respect HCLWattsUp (using a constant positive offset). The Euclidean distance will also be reduced from 92,104 to 26,502 after calibration. However, the calibration does not improve the overall qualitative difference in energy consumption behavior. The overall energy consumption trend remains different after calibrating RAPL readings. It suggests that the correlation coefficient, average error, and Euclidean distance are not sufficiently accurate measures for comparing the similarity of energy profiles.

The next case study demonstrates that a non-similar energy profile used as an input to an energy optimization algorithm can cause significant energy losses. In Figure 1, two sample energy profiles are compared against the ground truth (labeled *Real*). The average errors of profiles *Model1* and *Model2* against the ground truth are 62% and 64%. The Euclidean distance between profiles *Model1*, *Model2*, and the ground truth is 18,108 and 33,550. *Model1* and *Model2* are equally strongly correlated with the ground truth with the correlation coefficient equal to 0.91.

While *Model1* is ranked better than *Model2* by both the Euclidean distance and average error, it exhibits different energy consumption behavior for more than 40% of data points as compared with ground truth. Hence, it causes a significant loss of energy when input to the energy optimization algorithm [12], which employs the workload size as the decision variable for energy optimization of an application. For example, *Model1* only provides 21% of workload distributions that are the same as those provided by ground truth when used as an input to this algorithm for energy optimization. In contrast, *Model2* provides the same workload distributions as of the ground truth for 79% of problem sizes despite its higher average error and greater Euclidean distance. Therefore, *Model2* is better than *Model1* for the use in energy optimization or energy consumption analysis of the application.

Thus, the average error, Euclidean distance, and correlation coefficient are not sufficient to measure the similarity between energy profiles despite being the most used statistical measures for this purpose. The average error and Euclidean distance are highly sensitive to outliers and do not capture the similarity of energy consumption trends. They are also highly sensitive to the transformations such as uniform amplitude/time scaling, shifting, etc. Pearson correlation coefficient, on the other hand, assumes a linear relationship between the variables which might not be always true. It can also be easily misinterpreted as the high correlation coefficient does not necessarily mean a strong linear relationship or high similarity between two variables. Finally, they can mislead in many cases by erroneously grading the energy profile the best and causing significant energy losses when used for energy optimization of an application.

2.3. Trend-Based Similarity Measuring Methodology for Energy Profiles

We now present our solution method called trend-based similarity measure of energy profiles (TSM) to determine the similarity between the energy profiles and the ground truth. We use the term “model” to represent the regression model of the energy profile for illustration purposes, unless stated otherwise.

The inputs to TSM are the precision setting and a set of energy profiles (constructed with different energy measurement approaches) and the ground truth (EPS). The precision setting is the same as the experimental settings used to construct the energy profiles of the application. For example, for each data point in the energy function for an application, we repeatedly execute the application until the sample mean lies in the 95% confidence interval and a precision of 0.025 (2.5%) has been achieved. The output of TSM is the ranking of energy profiles based on their distance which reflects their resemblance with the ground truth.

TSM is composed of the following four stages: (i). The underlying regression model of the energy profile is learned, (ii). The regression fits of this energy profile and the ground truth are compared to determine if they exhibit the same trend, (iii). If they do not, then the energy profile is branded fundamentally inaccurate; (iv). If they do, the distance between the regression models of the energy profile (that follows the same trend-line as of the ground truth) and the ground truth is determined using Euclidean distance as a metric of goodness of the energy profile.

2.3.1. Model Fitting

Energy profiles are usually constructed as a function of problem size, CPU threads/cores, or CPU frequency. The configuration parameters have a strong influence on the overall energy consumption behavior of the application from the energy profiles. The experimental observation in our previous studies [13,14] is that the overall trend of the energy profile of an application is a monotonically increasing function of workload size. Like the energy profiles, their underlying energy consumption trend can be linear or non-linear, however, the general direction of energy consumption is monotonically increasing.

The authors in [11] also report the same. They find that the dynamic energy profiles of an application in the single-core era increase monotonically with problem size and are smooth linear functions of problem size. However, multicore CPUs exhibit inherent complexities including (a) non-uniform memory access (NUMA); (b) severe contention for shared on-chip resources such as last level cache

(LLC), interconnect, and DRAM controllers due to tight integration of tens of cores; and (c) dynamic power management (DPM) of CPU sockets and DRAM. Due to these complexities, while the trend of the energy profiles is still a monotonically increasing function of workload size, the functional shape is non-smooth and non-linear.

Therefore, the first step of TSM is based on the regression analysis of the energy profiles to examine their underlying regression models using the application configuration parameters as predictor variables to model the energy consumption. We use polynomial regression to model the relationship between the energy consumption of an application and its configuration parameter.

Polynomial regression fits the relationship between the dependent variable (the energy consumption) and the predictor variable (the application configuration parameter), as an n -th degree polynomial. Therefore, it can estimate both linear and non-linear models. For example, the linear models are fit as polynomial regression of degree 1 whereas the non-linear models are fit as higher (i.e., greater than 1) degrees of polynomial regression (such as quadratic, cubic, etc.). To facilitate clarity of exposition, the mathematical form of k th order of polynomial regression model [40] can be stated as follows:

$$f_E(y) = c_0 + \sum_{i=1}^n c_i x_i^k \quad (1)$$

where $x = \{x_1, \dots, x_n\}$ is the predictor variable; c_0 is the intercept; k is the degree of polynomial; and $c = \{c_1, c_2, \dots, c_n\}$ is the vector of coefficients (or the regression coefficients). In real life, there usually is stochastic noise (measurement errors). Therefore, the model can be expressed [41] as

$$\tilde{f}_E(y) = f_E(y) + \epsilon \quad (2)$$

where the error term or noise ϵ is a Gaussian random variable with expectation zero and variance σ^2 , written $\epsilon \sim \mathcal{N}(0, \sigma^2)$.

In [42], the authors report that the dynamic energy models having a non-zero intercept violate the basic principle of the theory of energy predictive models for computing. This is because dynamic energy is consumed by the CMOS component due to the switching activity when executing an application. There is no switching activity in the absence of workload execution, and therefore the system dissipates static energy only. Hence, regression models should predict no dynamic energy consumption when there is no workload. Therefore, to conform to this principle, we force the intercept to be zero while fitting the regression models.

To choose the order of the polynomial regression model reflecting the best fit, we follow a systematic approach called the forward selection procedure. In this approach, the models are successively fit in increasing order of polynomial and the significance of regression coefficients is tested at each step of model fitting. The order is kept increasing until an F-test for the highest order term is non-significant. Briefly, the F-test of overall significance indicates whether the regression model provides a better fit to the data than a model that contains no independent variables. It has the following two hypotheses: (i). The null hypothesis: It states that the model with no independent variables (intercepts only) fits the data equal to the regression model, and (ii). The alternative hypothesis: It states that the regression model fits the data better than the intercept-only model. A regression model is considered as significant if the p-value of the F-test is less than the significance level (i.e., 95% of the confidence interval or 0.05 level). We find that the first and third-order polynomials fit for all the energy profiles in our application suite.

We want to emphasize here that the purpose of fitting the regression model is not to build an offline energy model to predict the energy consumption by employing a predictor variable. Instead, the regression analysis is performed to examine the underlying model of energy profiles in an EPS to facilitate the comparison of their energy consumption trend. Therefore, we fit the regression model on the whole dataset instead of splitting it into training and test datasets.

2.3.2. The Discrepancy Analysis

In the second step, we compare the regression models of energy profiles in the EPS and the ground truth. We term this test as the discrepancy analysis. The following conditions must hold for the regression models of two energy profiles to be ideally alike:

1. The regression models of the energy profile and the ground truth must follow the same orientation. Both regression models must exhibit the same increase and decrease in the range of all the data points.
2. The regression models of the energy profile and the ground truth must not intersect at any point.
3. The distance between the regression models of the energy profile and the ground truth must be the same for the range of all the data points.

All these properties must be satisfied by a regression model to be considered as ideally similar to that of the ground truth.

Mathematically, the idea can be expressed as the slope and its direction must be the same for the regression models of an ideally similar energy profile and of the ground truth. The slope and its direction can be determined by taking the first and second derivatives of the regression models. Therefore, we compare the first and second derivatives of the regression equations of an energy profile and the ground truth. While the first derivative indicates whether the energy consumption trend is increasing or decreasing, the second derivative tells about the shape of the underlying regression model of the energy function. Two regression models do not follow the same trend if the second derivative of one of them is positive (greater than zero) and negative (less than zero) for the other one, or vice versa. However, they follow the same direction if the second derivatives of both regression models are either positive (greater than zero) or negative (less than zero). The regression models that do not follow the same direction are classified as *opposite* and consequently removed from the EPS.

In the next step, we compare the coefficients of the second derivatives of the regression models that follow the same direction. Two models are considered as *same* if the difference of coefficients of their derivatives is within an interval of input precision settings. To illustrate this, consider two third order (cubic) polynomial regression models r_1 and r_2 . Let r_1 is the regression model of the ground truth. Let the coefficients of the second derivatives of both models are c_1 and c_2 respectively. Then, the difference between the coefficients of the second derivatives of both models is calculated as $\epsilon = |(c_1 - c_2)/c_1 \times 100$. Now, if ϵ lies within the input precision settings, then both regression models are considered as the *same*. Otherwise, they are classified as *similar*.

To summarize, we analyze the qualitative behavior of regression models of the energy profiles and the ground truth by comparing the derivatives of their polynomial functions. As a result, we classify the energy models into one of the following three categories:

1. *Opposite*: The slopes of the regression models of an energy profile and the ground truth are in the opposite direction. The regression models of an energy profile and the ground truth exhibit opposite behavior such that one of them is increasing at x and the other one is decreasing with x . Furthermore, the shape of the regression fit is concave up for one of them and concave down for the other one.
2. *Same*: The slopes of the regression models of an energy profile and the ground truth are identically the same and follow the same direction. This class represents the energy profiles which are ideally the same to their corresponding ground truths.
3. *Similar*: The slopes of the regression models of an energy profile and the ground truth are different, however, they follow the same direction. It indicates that the energy profile is neither the same as the ground truth nor in the opposite direction to it.

As a result of this step, the energy profiles that have regression fits in the opposite direction to that of the ground truth are removed from the EPS. Consequently, the resulting EPS contains only the

same or similar energy models. The goodness of the remaining energy profiles to the ground truth is quantified in the third step.

2.3.3. The Distance Metric

In this step, we determine the distance between the regression fit of each energy profile and the ground truth in the remaining EPS. For this purpose, we use Euclidean distance as a distance metric to establish an absolute value of the distance of the regression fit of each profile and the ground truth. Because of its triangular-inequality property, Euclidean distance is used to index the model space which speeds up the search and matching in general, especially for the huge model space. Furthermore, it helps in ranking the similar profiles based on their distance with the ground truth in an EMS.

Hence, to rank the profiles based on their similarity, first, the Euclidean distance between the regression models of the profiles and the ground truth belong to the same EPS is computed. Then, the energy profiles are ranked according to the Euclidean distance between their regression models and that of the ground truth. The energy profile whose regression model has the least Euclidean distance with that of the ground truth is considered as the most similar profile in that EPS. The final output of this step is the sets of the energy profiles with similarity ranks. Two profiles may have the same rank if their Euclidean distance differs by less than or equal to the input precision.

2.4. Experimental Setup

In this section, we explain the experimental platform, the application dataset used in this work, and the experiment methodology used to validate TSM.

2.4.1. Experimental Platform and Applications

The dataset used in this work comprises of 235 energy profiles of different application configurations executed on multicore heterogeneous hybrid computing platforms and constructed with on-chip sensors, power meters, or energy predictive models employing performance monitoring counters (PMCs) as predictor variables. The profiles are constructed as the results of the research works [13,14]. The details on experimental setup, platforms, application suite, configuration parameters, and the boundary conditions to construct the dataset are presented in the Supplemental Materials [37] of this work. Briefly here, the application configuration parameters are (a) problem size, (b) number of CPU threads, or the number of CPU cores. The application suite used to construct the profiles contains highly optimized memory bound and compute-bound scientific routines and two unoptimized routines. The optimized routines include matrix multiplication employing DGEMM offered by OpenBLAS package, matrix multiplication employing DGEMM offered by Intel Math Kernel Library (MKL), two-dimensional FFT (2D FFT) from FFTW package, 2D FFT from Intel MKL, benchmarks from NASA Application Suite (NAS), high-performance conjugate gradient (HPCG) from Intel MKL, and stress. The unoptimized routines are basic matrix multiplication and matrix-vector multiplication.

We employ three nodes for constructing the energy profiles. The first node is HCLServer01. It has an Intel Haswell E5-2670 multicore CPU containing 24 physical cores with 64 GB DDR4 main memory. It hosts two accelerators, one Nvidia K40c GPU and one Intel Xeon Phi 3120P. The Nvidia GPU has 3584 processor cores with 12 GB main memory and memory bandwidth of 549 GB/s. The Intel Xeon Phi contains 57 processor cores with 6 GB GDDR5 main memory and memory bandwidth of 240 GB/s. The second node is HCLServer02. It contains an Intel Xeon Gold 6152 Skylake multicore CPU consisting of 22 cores and 96 GB DDR4 main memory. It hosts one Nvidia P100 GPU. The GPU has 2880 processor cores with 12 GB GDDR5 main memory and memory bandwidth of 288 GB/s. The third node is HCLServer03. It hosts an Intel Xeon Platinum 8180 Skylake multicore CPU having 56 cores with 187 GB main memory.

A Watts Up Pro power meter is installed between the wall A/C outlet and the input power socket of a node. The power meters are calibrated periodically using a Yokogawa WT310 power meter, which is an ANSI C12.20 revenue-grade power meter. The sampling speed of Watts Up Pro power

meters is one sample every second. The datasheet reported accuracy is $\pm 3\%$. The unit of measurement is 0.5 watts.

For the on-chip sensor measurements, we use RAPL [38] to determine the energy consumed by the application kernels executing on Intel CPUs. For the Nvidia GPUs, NVIDIA Management Library (NVML) [43] is employed. And for Intel Xeon Phi, the Intel System Management Controller chip (SMC) [44] (using Intel manycore platform software stack (MPSS) [45]) is utilized. HCLWattsUp interface [39] is used to obtain the power measurements from the WattsUp Pro power meters. The Intel MKL installed in the nodes has version 2017.0.2; the CUDA versions present on HCLServer01 and HCLServer02 are 7.5 and 9.2.148.

The statistical methodology used to obtain a data point reliably using the different tools is explained in [37]. Briefly, the methodology determines the sample mean (execution time or dynamic energy or PMC) by executing the application in a loop. The loop is terminated when the sample mean meets the statistical confidence criteria (95% confidence interval, precision of 0.025 (2.5%)). The student's *t*-test is employed to determine the sample mean. Pearson's chi-squared test is used to ensure that the observations follow a normal distribution to satisfy the assumptions of the test.

2.4.2. Experimental Methodology to Validate TSM

We classify our suite of energy profiles (EPS) into the following two groups:

1. Group A (Sets of many energy profiles): Group A comprises of the EPS where there is more than one energy profile of the same application constructed with different approaches such as on-chip power sensors, system-level power measurements provided by power meters, etc.
2. Group B (Sets of single energy profiles): Group B comprises of the EPS where only one energy profile is compared with the ground truth.

For each group, we fit the regression models as *n*-th order of polynomial for each energy profile and their corresponding ground truths belonging to the same EPS. To choose the best order of polynomial approximation, we follow the forward selection procedure as explained in Section 2.3.1. Intuitively, the polynomial order should be the same for the regression models of an energy model and its corresponding model of the ground truth belonging to the same EPS. As a result of this sanity check, we reject the energy functions which have a different order of polynomial as a best fit than the regression model of the ground truth.

In the next step, we analyze the qualitative behavior of regression models of the energy profiles and the ground truth by comparing the derivatives of their polynomial functions as explained in Section 2.3.2. The energy profiles classified as opposite are removed from their respective EPS, as a result of this step. In the third step, we determine the similarity between the remaining energy profiles and the ground truth using Euclidean distance. We compare the results of TSM with other statistical approaches such as correlation, Euclidean distance, and average error to compare the accuracy and similarity of energy profiles (and for the time series of equal length in general). The Euclidean distance that is compared with TSM is the distance between the energy profiles and the ground truth. In contrast, TSM uses the Euclidean distance between the regression models of the energy profiles and the ground truth in an EPS.

3. Results and Discussion

This section is structured as follows: (i). Comparison of the accuracy of energy profiles determined with TSM and other popular statistical approaches, (ii). A general discussion on the results obtained and their interpretation, and finally (iii). Comparison of the effectiveness of TSM with other popular statistical approaches using a profile-based energy optimization algorithm as a yardstick that employs the workload size as a decision variable.

3.1. Experiment Results

Group A (Sets of many energy profiles): The similarity ranking by TSM and popular statistical approaches for the energy profiles in each EPS that belongs to group A are provided in the supplemental [37]. One can observe that the correlation coefficient does not always distinguish much between the energy profiles. Consider, for example, the profiles in EPS DGEMM_EqualLoad. The profiles RAPL_Parallel and RAPL_Combined both have a correlation of positive 0.9993 with the ground truth (HCLWattsUp_Parallel). Similarly, the correlation coefficient for RF_Additive and NN_Additive is 0.9999 with the ground truth in EPS DGEMM_Predictive Models.

Similarly, the average prediction error also misleads in many cases. Consider, for example, the profiles in EPS DGEMM_EqualLoad. The average prediction error suggests the HCLWattsUp_Combined as the most similar profile with the ground truth. However, TSM suggests RAPL_Combined as the most similar, and HCLWattsUp_Combined as the most different among all three profiles in EPS DGEMM_EqualLoad. A visual illustration of regression models of the profiles in EPS DGEMM_EqualLoad as presented in Figure 3, also conforms to the TSM. In general, one can observe that regression models of all three profiles in EPS DGEMM_EqualLoad follow the same pattern as of the ground truth (HCLWattsUp_Parallel). However, both the RAPL_Parallel and RAPL_Combined exhibit the closest resembling pattern with the ground truth for the range of all problem sizes as illustrated in Figure 3b. HCLWattsUp_Combined, on the other hand, exhibits a slightly different pattern at both ends of data points (that is the range of very small problem sizes and very large problem sizes) as shown in Figure 3a. Therefore, while it follows the same orientation, it is ranked as the least similar profile in its EPS.

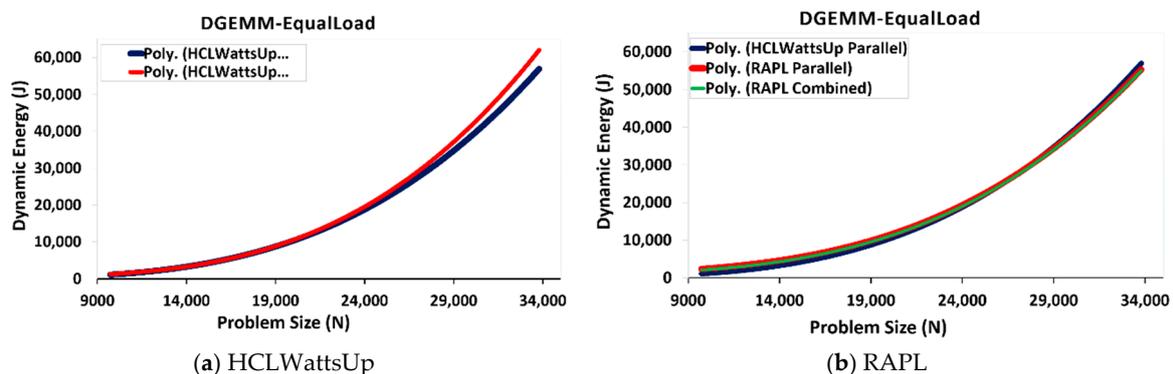


Figure 3. Regression models of energy profiles in set of energy profiles (EPS) DGEMM_EqualLoad constructed with (a) HCLWattsUp, and (b) RAPL. Here, DGEMM refers to Double-precision General Matrix Multiplication application.

However, the similarity results as presented in the supplemental [37] suggest that overall the Euclidean distance between the energy profiles proves to be more efficient than the correlation coefficient and average prediction error. In most of the cases, it suggests the similarity ranking in line with TSM. However, it also misleads in some of the cases. Consider, for example, the similarity ranking for the profiles in EPS DGEMM_Predictive Models. Euclidean distance between the profiles ranks LM_additive as the third most similar profile, whereas TSM ranks it as the fifth most similar profile. Similarly, Euclidean distance between the profiles ranks RF_NonAdditive as the second most similar in EPS FFT_Predictive Models, whereas TSM ranks it as the third most like the ground truth in its EPS.

It is important to note that the statistical measurements and the metrics do not capture the holistic picture of the energy consumption trend of the profiles. Consider, for example, the profiles in EPS DGEMM_AnMoHA. Both the Euclidean distance and average prediction error consider the profiles Combined_3 and Combined_4 as the third most similar and fourth-most similar profiles with the ground truth (Parallel). However, one can observe in Figure 4b that the qualitative comparison of the regression fit of both profiles and the ground truth by TSM suggests them to have a different

energy consumption trend and thus drop them from the EPS. The correlation coefficient ranks the profiles in this EPS in line with TSM and ranks them as least similar. But it also does not provide the details on their qualitative difference of the underlying energy consumption behavior. Similarly, both the Euclidean distance and average prediction error rank Combined_2 as the least similar profile. In contrast, TSM ranks it as the third most similar profile.

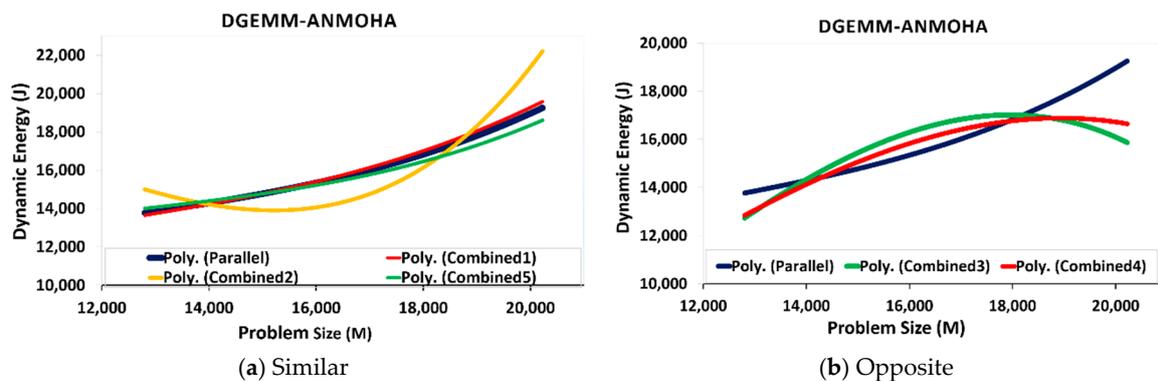


Figure 4. Regression models of energy profiles in set of energy profiles (EPS) DGEMM_AnMoHA. (a) Combined_1 and Combined_5 follows the same trend, Combined_2 follows similar trend, and (b) all profiles exhibit opposite trend as of the ground truth. Here, DGEMM refers to Double-precision General Matrix Multiplication application, and AnMoHA refers to Additive Energy Modelling of Hybrid Applications.

Graphical illustration of regression models of the profiles in EPS DGEMM_AnMoHA as presented in Figure 4 also confirms the same results. One can observe that Combined_3 and Combined_4 exhibit different energy consumption behavior as of the ground truth (Parallel). However, Combined_1 and Combined_5 follow the same direction with the same slope, whereas Combined_2 follows the same direction but exhibits a different orientation.

Group B (sets of single energy profile): For group B, TSM classifies the similarity of energy profiles with ground truths (after comparing their regression fits) into the three similarity categories explained in Section 2.3.2. The similarity ranking by TSM and popular statistical approaches for the energy profiles in each EPS that belongs to group B are provided in the supplemental [37]. One can observe that likewise group A, all three statistical approaches fail to capture the qualitative difference of the regression models of the energy profiles and the ground truth belong to the same EPS.

Consider, for example, the regression models of the energy profiles illustrated in Figures 5 and 6 representing the classes *same* and *similar* respectively. The regression models of the energy profiles follow the same trend as the ground truths in both cases. However, the slopes of the regression models presented in Figure 6 are different from their corresponding ground truths. Figure 7 illustrates the regression models representing the class *opposite*. It can be observed that the regression models of the energy profiles and their corresponding ground truths exhibit different trends. Consider, for example, the regression models of the EPS {FFTW,G = 16,T = 7}. Here, G and T represent the number of thread groups and the number of threads per group respectively. The slopes of the regression models of both profiles are different and have different signs, positive for RAPL and negative for HCLWattsUp. Figure 7c also shows the same results. One can observe while the shape of the regression model of RAPL is concave up, it is concave down for HCLWattsUp. However, the popular statistical approaches do not capture this behavior.

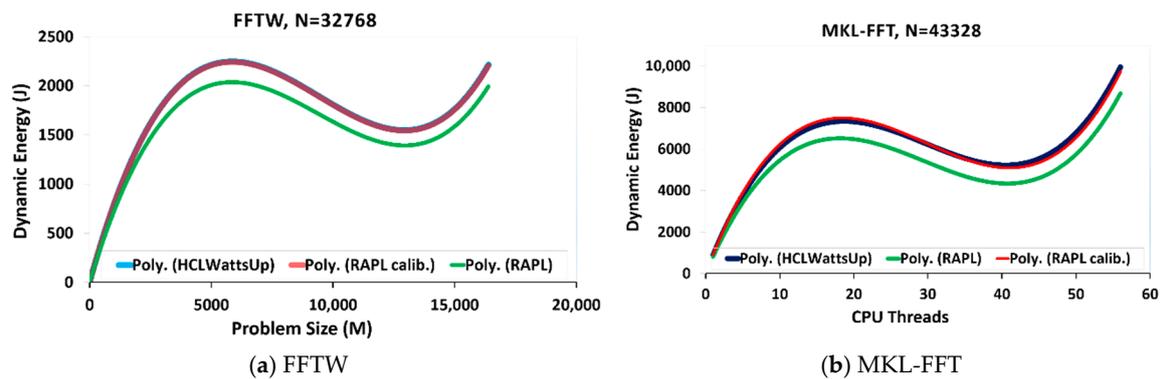


Figure 5. Group B (sets of single energy profile), class same similarity. Group B comprises of sets of energy profiles where only one energy profile is compared with the ground truth. Here, FFTW refers to Fastest Fourier Transform in the West, and MKL-FFT refers to Intel Math Kernel Library kernel for Fastest Fourier Transform.

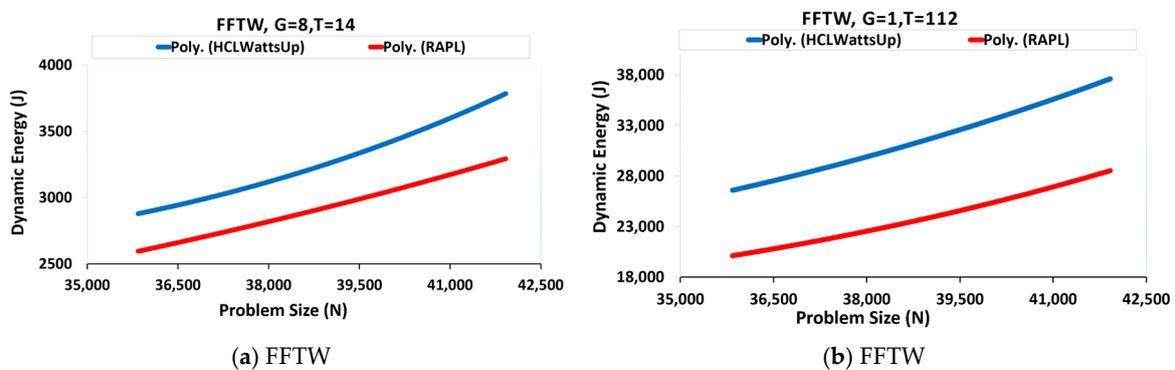


Figure 6. Group B (sets of single energy profile), class similar similarity. Here, G = CPU thread groups, and T = CPU threads.

3.2. Discussion

Average error and Euclidean distance do not indicate whether the calibration can improve the average error or Euclidean distance between two similar energy profiles, and thus can mislead to consider an accurate energy profile as inaccurate. However, unlike average prediction error and Euclidean distance, TSM can indicate if the average prediction error and Euclidean distance can be reduced by calibrating the energy profile with the ground truth in an EPS. Consider, for example, the energy profiles in EPS FFTW where the configuration parameter is the problem size $M \times N$ where $M \leq N$, and $N = 32,768$. Figure 5a illustrates the regression models of the profiles. The difference between the slopes of the regression fit of the RAPL energy profile and the ground truth is very close to zero, and thus TSM classifies their similarity as the *same*. This *same* similarity suggests that the regression models of both the RAPL energy profile and the ground truth exhibit the same energy consumption behavior. Therefore, one can reduce the average prediction error and Euclidean distance between the RAPL energy profile and the ground truth from 10.5% to 0.6% and from 1134.9 to 94, respectively, after calibrating it with the ground truth. That is an improvement of 94% in average prediction error and 92% in the Euclidean distance between the profile.

Similarly, consider the EPS IntelMKLFFT where the configuration parameter is CPU cores and problem size N is 43,328. Figure 5b illustrates the regression models of the profiles. The difference between the slopes of the regression models of both the RAPL energy profile is close to zero, but slightly more than the difference between the energy profiles belong to the EPS FFTW. TSM classifies both the RAPL energy profile and the ground truth as *same*. After calibration, one can reduce the average prediction error and the Euclidean distance between the profiles from 13% to 2.19% and from

6700 to 1495.83 respectively. That is an improvement of 83% in average prediction error and 78% in Euclidean distance between the RAPL energy profile and the ground truth in that EPS.

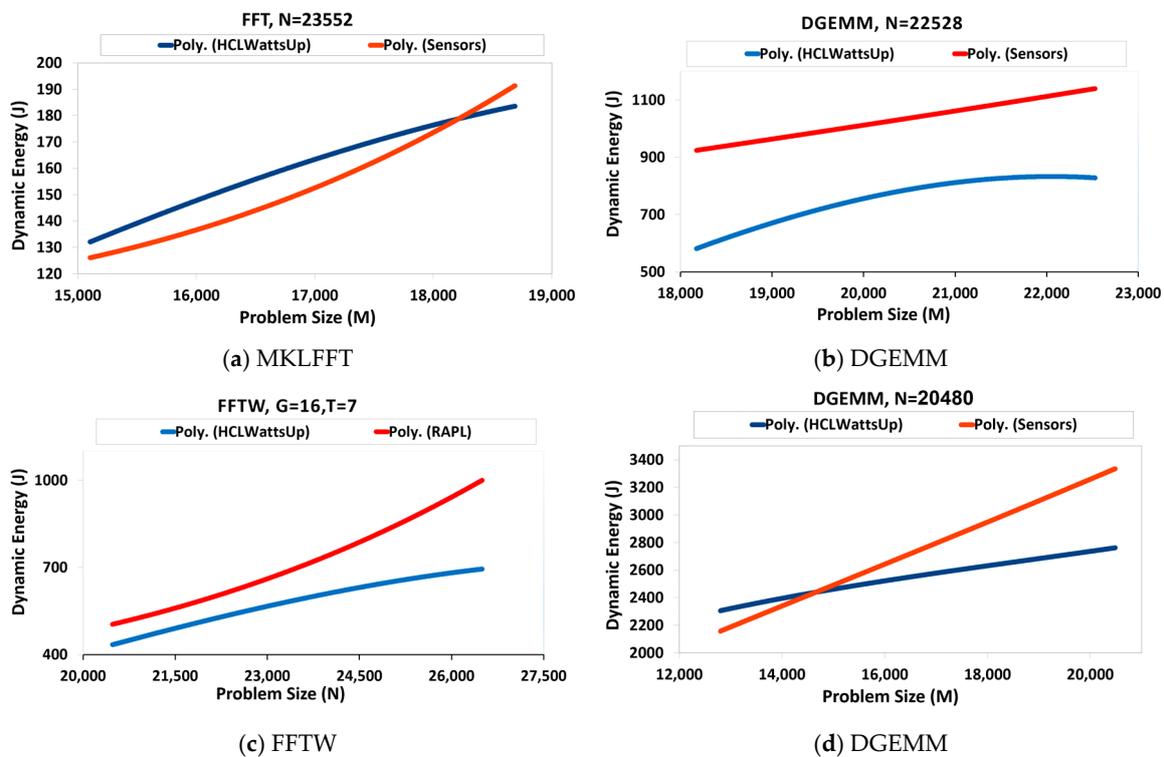


Figure 7. Group B (sets of single energy profile), class opposite similarity. Here G = CPU thread groups and T = CPU threads. Here, FFT refers to Fastest Fourier Transform; FFTW refers to Fastest Fourier Transform in the West; and DGEMM refers to Double-precision General Matrix Multiplication.

It is important to note while the similarity classes such as *opposite* and *same* are more useful for group B, the similarity class *similar* provides less information. It does not present any threshold to indicate the absolute value of the similarity between the energy profile and the ground truth. The threshold that indicates the value of absolute similarity is dependent on the application domain.

Consider, for example, signal processing or multimedia processing applications, which are considered as fault-tolerant and belonging to the approximate computing domains. An inaccurate result is acceptable in such domains. Therefore, a comparatively less similar energy profile can also serve the purpose in this case. In contrast, a high similarity value is required for applications such as cryptography or hard real-time applications. That is why TSM does not define a threshold to indicate the degree to which an energy profile exhibits a similar energy consumption behavior to the ground truth. Instead, it just compares the energy consumption behavior and the shapes of the regression models of the energy profile with the ground truth in an EPS and determines whether both have similar shape and energy consumption behavior.

To quantify the *similar* energy profiles, one can take the difference of the polynomials or the derivatives of the regression models of the energy profile under consideration and the ground truth. A zero value of the difference between the polynomials or derivatives indicates the same polynomials and thus the same regression models. One can give some weight to the energy profile under consideration indicating how large is it from zero value of the difference, and thus how less similar is it with the ground truth.

Unlike the profiles classified as *same*, there is little to none margin for average error and the Euclidean distance reduction, if the profile is classified as *similar*, after calibration. This is because the derivatives of the regression models of the energy profile under consideration and the ground truth

have different slopes. Therefore, the calibration can reduce the average error and Euclidean distance between the energy profile and the ground truth only to an extent. However, it highly depends on the value of the similarity between the polynomials/derivatives of the regression models of the energy profile and the ground truth in an EPS.

Consider, for example, the EPS FFTW where problem size N ranges from 35,840 to 41,920 and the configuration of CPU threads are grouped into 8 and there are 14 CPU threads in each group. We refer to this EPS as EPS_1 for illustration purposes. Figure 6a illustrates the regression models of the profiles in EPS_1 . One can reduce the average prediction error and Euclidean distance between the RAPL energy profile and the ground truth from 13.66% to 12.73% and from 5569.4 to 4520.9 after calibration. That is an improvement of 6.81% in average prediction error and 18% in Euclidean distance between the RAPL energy profile and the ground truth in EPS_1 . In contrast, consider the EPS FFTW where problem size N ranges from 35,840 to 41,920 and all 112 CPU threads are grouped into 1 group. Let this EPS be EPS_2 for illustration purposes. Figure 6a illustrates the regression models of the profiles in EPS_2 . The average prediction error and Euclidean distance between the RAPL energy profile and the ground truth can be reduced from 24.62% to 3.9% and from 78,669 to 23,184.7 after calibration. That is an improvement of 84.16% in average prediction error and 70.5% in Euclidean distance between the RAPL energy profile and the ground truth in EPS_2 . This is because the difference in polynomials and derivatives of the regression models of both profiles in EPS_2 is less than the regression models in EPS_1 .

Another important finding is that the calibration of less similar profiles with the ground truth can increase the maximum prediction error between them in some cases when trying to reduce the average error and Euclidean distance. Consider, for example, the EPS EPS_1 . The maximum prediction error between the RAPL energy profile and the ground truth is 29.8% which increases to 55.65% after calibration (using the same offset that reduces the average prediction error and Euclidean distance). That is an increase of 87% in the maximum error. We observe similar findings with other less similar energy profiles. However, it is not the case where the similarity is higher or the *same* between the energy profile under consideration and the ground truth in an EPS. The calibration improves the maximum error, average prediction error, and Euclidean distance between such profiles.

The calibration should only be applied to profiles that exhibit a similar energy consumption trend as of the ground truth because it only improves the Euclidean distance and prediction error, and thus, does not improve the qualitative difference of the energy consumption trend of the profiles.

TSM also indicates whether the predictive model which is employed to construct the energy profile, includes some extraneous contributor that does not reflect the energy consumption by the application, or it lacks some essential contributor to the energy consumption by the application. Consider, for example, the similarity results for EPS FFT_Predictive Models as presented in the supplemental [37]. The profile LM_NonAdditive has the highest average error and the greatest Euclidean distance with the ground truth. However, the slopes of the regression models of LM_NonAdditive are in the same direction as the ground truth (HCLWattsUp). Furthermore, the difference between its polynomials and the slopes is close to that of the LM_Additive. However, LM_NonAdditive predicts energy consumption more than the ground truth and LM_Additive. It suggests that the predictive model of LM_NonAdditive includes some extraneous PMC which does not reflect the energy consumption by the application.

Therefore, we apply a constant negative offset to its predictions to calibrate them with ground truth. As a result of this calibration, the average error and Euclidean distance of LM_NonAdditive energy profile with the ground truth is reduced from 92% to 39%, and from 3321 to 2722. This is an improvement of 58% in average error and 18% in Euclidean distance. Consequently, the calibrated energy profile of LM_NonAdditive is closer to LM_Additive in terms of its average error and Euclidean distance with ground truth. One can observe in Figure 8 that the regression model of calibrated LM_NonAdditive is in a closer approximation of the ground truth (HCLWattsUp) and the profile LM_Additive.

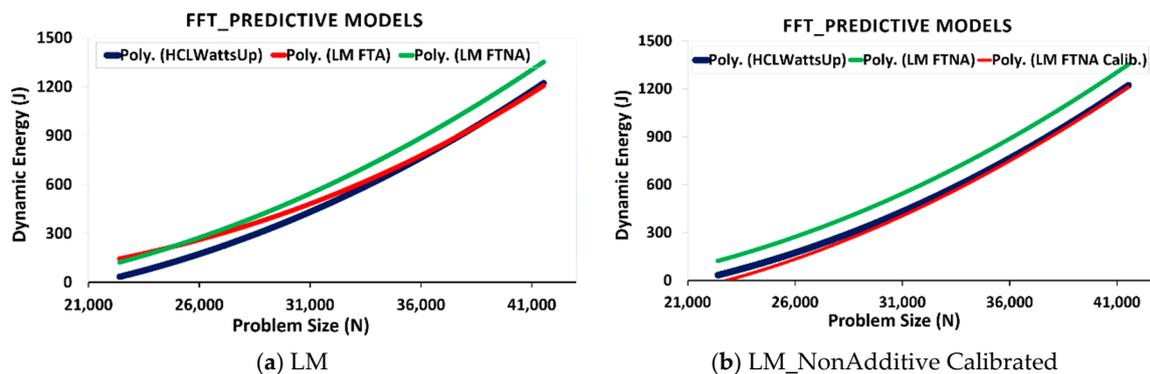


Figure 8. Regression models of energy profiles in EPS FFT_Predictive Models, (a) Linear Models, and (b) Linear Model NonAdditive Calibrated. Here, LM refers to Linear Model.

This suggests that the prediction error of the energy profile LM_Additive can be improved by removing non-relevant PMCs from the set of explanatory variables. This finding conforms to the results as presented in [42], where the authors present a study to demonstrate how the prediction errors of PMC based energy predictive models can be improved significantly by removing irrelevant PMCs (which does not reflect the energy consumption by the application) from the set of predictor variables.

Similar to the indication of extraneous PMCs, TSM can also indicate whether the predictive model, which is employed to construct the energy profile, lacks essential PMCs which strongly reflects the energy consumption by the application. Consider, for example, the regression models of energy profiles of FFTW_32768 and MKLFFT_43328 as shown in Figure 5. The energy profiles of the application with RAPL exhibit the *same* energy consumption patterns as of the ground truth. However, RAPL under-reports energy consumption in comparison with the ground truth. It suggests that the energy profiles of both applications lack the contributions by some essential components. The prediction errors and Euclidean distance of both profiles can be reduced significantly by applying a constant positive offset to its predictions to calibrate them with the ground truth. The calibration improves the average prediction error and Euclidean distance of FFTW_32768 by 94% and 92%, respectively, and by 83% and 78%, respectively, for MKLFFT_43328. Hence, TSM may be used as a selection criterion of PMCs in energy predictive models to predict the energy consumption by an application. We will investigate this direction in our future work.

To summarize, the statistical approaches (correlation coefficient, average prediction error, and the Euclidean distance between the energy profiles) fail to distinguish the energy profiles based on their underlying energy consumption trend. They erroneously rank an inaccurate energy profile as better than more accurate ones in some cases. TSM, on the other hand, proves to be more effective in capturing the energy consumption behavior of the profiles and comparing their qualitative differences. It provides more information about the energy consumption behavior of the profiles and thus ranks them based on their proximity with energy consumption behavior of the ground truth. Furthermore, it can also suggest if the calibration can improve the Euclidean distance, and the average and maximum prediction errors between the energy profile under consideration and the ground truth.

3.3. Comparison of TSM and State-of-the-Art Statistical Approaches for Energy Optimization

In this section, we compare the effectiveness of TSM with other popular statistical approaches using a profile-based energy optimization algorithm as a yardstick that employs the workload size as a decision variable. Furthermore, we demonstrate that inaccurate energy profiles can cause a significant amount of energy loss when used for the optimization of an application for dynamic energy.

The profile-based energy optimization algorithms [11,12] leverage the variations (jumps and drops) of the energy profiles and determine the workload distributions that optimize the total dynamic energy consumption for the given workload size. These variations in energy profiles are caused by the intrinsic complexities in modern hybrid heterogeneous computing platforms such as resource

contention due to non-uniform memory access (NUMA) and the tight integration of multi-core CPU with one or more accelerators. The algorithm provides the same workload distributions for energy profiles classified as identically the same (second stage of TSM) used as an input for the range of the same workload sizes. Likewise, it provides different workload distributions for non-similar energy profiles used as input for the range of the same workload sizes. This is because non-similar energy profiles exhibit different variations in their energy consumption behavior for the same set of data points.

We use the profile-based optimization method [12] to determine the optimal partitioning of the workload size to optimize the total dynamic energy consumption of the application. The energy optimization algorithm does not make any assumptions about the shape of input energy profiles. The algorithm takes the following inputs: (i). The workload size, (ii). The number of processors, and (iii). The discrete dynamic energy functions of individual processors. The output is the optimal workload distribution that provides minimal dynamic energy consumption for the input workload size. The algorithm has a polynomial complexity of $O(m^3 \times p^3)$. We compare the output workload distributions provided by the algorithm when using as an input the dynamic energy profiles ranked as similar to ground truth by popular statistical approaches and TSM.

For our first case study, consider the profiles in the EPS, DGEMM_AnMoHA as illustrated in Figure 9. The combined energy profiles are constructed following the additive energy modelling approach as presented in [13]. Briefly, the approach is based on the hypothesis that the total dynamic energy consumption during an application execution will be equal to the sum of energies consumed by all the individual application components executing on processors in the case of loosely-coupled application components. Formally speaking, let $E_A(x)$, $E_B(x)$, and $E_C(x)$ be the dynamic energy consumptions by the application kernels of workload size x executing sequentially on processors CPU1, GPU1, and PHI1. Let $Combined_{ABC}(x)$ represent the sum, $E_A(x) + E_B(x) + E_C(x)$. Let $Parallel_{ABC}(x)$ be the total dynamic energy consumption by parallel execution of the same application kernels of the workload size x on the processors CPU1, GPU1, and PHI1. Then, the additive hypothesis holds only if $Parallel_{ABC}(x) = Combined_{ABC}(x)$.

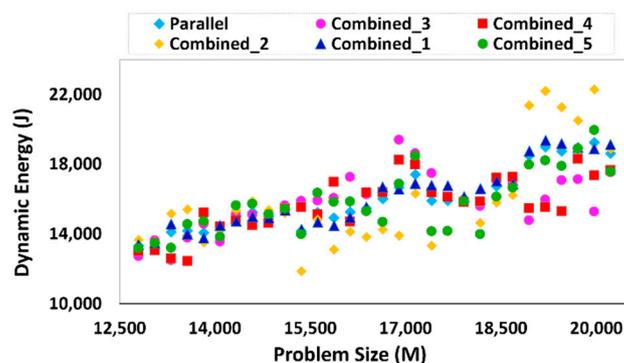


Figure 9. Dynamic energy profiles in set of energy profiles (EPS) DGEMM_AnMoHA.

We run a parallel hybrid application DGEMM (which multiplies two dense matrices A and B of sizes $M \times N$ where $M \leq N$) as explained in [13] on HCLServer01 for the workload sizes ranging from $38,400 \times 20,224$ to $60,672 \times 20,224$ with a constant step size of 256. The dimension M is equally partitioned among three aforementioned processors (CPU1, GPU1, PHI1) into M_1 , M_2 and M_3 such that the matrix $M_1 \times N$, $M_2 \times N$ and $M_3 \times N$ (i.e. $12,800 \times 20,224$) are computed by processor CPU1, GPU1, and PHI1 respectively. There are no communications involved between the processors. The DGEMM energy profiles in DGEMM_AnMoHA are constructed using different combinations of additive models of application-components executing on processors. More details on additive energy modelling of hybrid parallel applications and the design configurations of independent experiments to construct the energy profiles in DGEMM_AnMoHA can be found in [13].

Figure 9 illustrates the energy profiles in DGEMM_AnMoHA. The average prediction error, correlation coefficients and Euclidean distance of all energy profiles are {2%,8%,7%,6%,4%}, {0.9762,0.8641,0.5741,0.6741,0.8945} and {2258,8795,8421,7523,4515} respectively as presented in the supplemental [37]. The Euclidean distance and average prediction error rank the profiles Combined_3 and Combined_4 as the most similar to the ground truth (Parallel) after Combined_1 and Combined_5. However, one can observe in Figure 9 that the qualitative comparison of both profiles with the ground truth by TSM suggests them to have a different energy consumption trend and thus drop them from the EPS. The correlation coefficient ranks the profiles in this EPS in line with TSM and ranks them as the least similar. But it does not provide the details on their qualitative difference such as the underlying energy consumption trend of the profiles. Similarly, both the Euclidean distance and average prediction error ranks Combined_2 as the least similar profile in its EPS. In contrast, TSM ranks it as the third most similar profile. However, all three statistical approaches likewise TSM rank Combined_1 as the most similar energy profile.

We determine the workload distributions for workload sizes ranging from $38,400 \times 20,224$ to $60,672 \times 20,224$ using the individual additive dynamic energy profiles of each processor CPU1, GPU1, and PHI1 as an input to the data partitioning algorithm [12]. Combined_2 provides 32% of the workload distributions the same as of Combined_1 whereas Combined_3 and Combined_4 provide 29% and 20% same workload distributions as of Combined_1. This conforms to the results of TSM, which ranks Combined_2 as better than Combined_3 and Combined_5.

Workload distributions when using Combined_3 and Combined_5 consume more dynamic energy for 82% and 81% of the data points, respectively, in comparison with Combined_2 for the workload sizes. Consider, for example, the workload sizes {47616,48128,49664,50176,50688,51200,51712}. The workload distributions of Combined_3 consume {52%,51%,52%,51%,54%,49%,50%} respectively more dynamic energy than the corresponding workload distributions of Combined_2 for the workload sizes. Similarly, consider the workload sizes {47104,47616,48128,49152,49664,50176,51200}. The workload distributions of Combined_4 consume {39%,38%,40%,37%,36%,38%,40%} respectively more dynamic energy than the corresponding workload distributions of Combined_2 for the workload sizes.

For our next case study, consider the profiles in the EPS, DGEMM_EqualLoad in Figure 10. The energy profiles of DGEMM in this EPS are constructed with RAPL and HCLWattsUp when running equal workload sizes on each CPU socket of a dual-socket multi-core Intel Haswell platform (technical specifications are provided in the supplemental [37]). The details on energy profiles and their construction procedure can be found in [13]. Briefly, we equally partition the workload sizes ($M \times N$) ranging from $19,456 \times 9728$ to $67,584 \times 33,792$ on both CPU sockets such that the matrix $M_1 \times N$ and $M_2 \times N$ are computed by processor CPU socket1 and CPU socket2 respectively. There are no communications involved between the processors. Figure 10 shows the parallel and combined dynamic energy profiles of both application configurations.

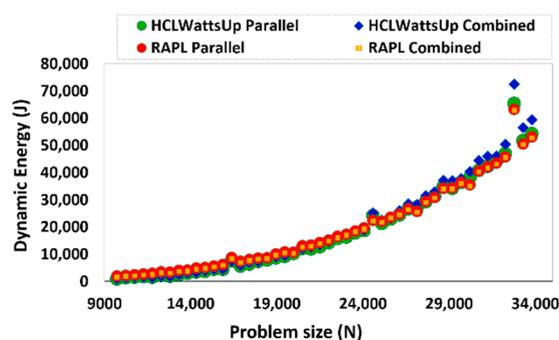


Figure 10. Dynamic energy profiles of DGEMM application in the set of energy profiles (EPS) DGEMM_EqualLoad.

All three energy profiles have almost the same strong positive correlation with the ground truth. The average errors of HCLWattsUp_Combined, RAPL_Parallel, and RAPL_Combined with HCLWattsUp_Parallel are 4.6%, 21.2%, and 16.1%, respectively. The correlation coefficient is the same (0.9993) for both RAPL_Parallel and RAPL_Combined, and 0.9995 for HCLWattsUp_Combined. Hence, both the correlation coefficient and average prediction error ranks HCLWattsUp_Combined as the most accurate energy profile in its EPS. However, TSM ranks it as the least similar. It ranks, in contrast, RAPL_Combined as the most similar to the ground truth in that EPS.

We determine the workload distributions for workload sizes ranging from $19,456 \times 9728$ to $67,584 \times 33,792$ using the dynamic energy profiles constructed with RAPL and HCLWattsUp as an input to the data partitioning algorithm [12]. Using the workload distribution, we run the applications in parallel on both sockets and determine its dynamic energy consumption with RAPL and HCLWattsUp separately. We find that the workload distributions when using HCLWattsUp_Combined consuming more dynamic energy for 65% of the data points of the range. Consider, for example, the workload sizes {56320,56832,57344,57856,58368,58880,59392,59904,60928}. The total dynamic energy losses by using HCLWattsUp_Combined in comparison with RAPL_Combined to optimize the dynamic energy consumption of DGEMM for the aforementioned workload sizes is {17%,18%,18%,17%,18%,18%,18%,18%,17%}, respectively.

To summarize, we use an energy optimization algorithm as a yardstick to evaluate the effectiveness of TSM and popular statistical approaches to be used in an energy optimization loop of an application. In all the presented case scenarios, TSM proves to be more effective. The energy profiles ranked as similar by TSM provide a greater number of same workload distributions as of the ground truth when using as an input to the energy optimization algorithm. Another important finding is that the energy profiles erroneously ranked as similar by popular statistical approaches can cause a significant amount of energy loss when used for the energy optimization of the application.

4. Conclusions

In this work, we presented a novel similarity measuring technique which considers the underlying energy consumption trend of the energy profiles. The proposed method captures the qualitative differences of the energy consumption behavior of energy profiles and ranks them based on their similarity with the ground truth. It effectively addresses the challenge of determining the goodness of application energy profiles on multicore computing nodes omnipresent in cloud infrastructures, supercomputers, data centers, and heterogeneous computing clusters where the shapes of energy profiles are non-smooth and non-linear. We compared the proposed method with popular statistical approaches, which are used to estimate the similarity between energy profiles, for a diverse set of 235 energy profiles (constructed on multicore heterogeneous hybrid computing platforms using state-of-the-art energy measurement techniques such as integrated power sensors, external power meters, or energy predictive models using PMCs as predictor variables). We demonstrated that the use of the state-of-the-art similarity approaches instead of the proposed one in the energy optimization loop leads to significant energy losses (up to 54% in our case).

We also showed that the proposed method can help determine whether the prediction model (that is employed to construct the profile) includes some extraneous contributor that does not reflect the energy consumption by the application or lacks some essential contributor to the energy consumption by the application. This finding further helps in determining whether the calibration can improve the average and maximum errors, and Euclidean distance between the energy profiles (constructed with over-estimated or under-estimated energy measurements) and the ground truth. Future work would include studying the efficiency of the proposed solution method in selecting the predictive model variables such as PMCs in order to improve their prediction accuracy.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1996-1073/13/15/3944/s1>.

Author Contributions: Conceptualization, M.F., A.S., R.R.M., and A.L.; methodology, M.F. and A.L.; software, M.F.; validation, M.F. and A.S.; formal analysis, M.F. and A.L.; investigation, M.F.; resources, M.F., A.S. and R.R.M.; data curation, M.F., A.S. and R.R.M.; writing—original draft preparation, M.F.; writing—review and editing, A.S., R.R.M. and A.L.; visualization, M.F., A.S. and R.R.M.; supervision, R.R.M.; project administration, A.L.; funding acquisition, A.L. All authors have read and agreed to the published version of the manuscript.

Funding: This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number 14/IA/2474.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this work:

IEA	International Energy Agency
ICT	Information and communication technology
TWh	Terra-watt hours
GPU	Graphics processing unit
FPGA	Field programmable gate array
QPI	Quick path interconnect
NUMA	Non-uniform memory access
DGEMM	Double-precision general matrix multiplication
FFT	Fast fourier transform
MKL	Intel Math Kernel Library
RAPL	Running average power limit
NVML	NVIDIA Management Library
HPC	High performance computing
TSM	Trend-based similarity measure
EPS	Set of energy profiles of an application constructed with different energy measurement approaches

The following nomenclature and units are used in this work.

Variable	Unit
Power	Watt
Static power	Watt
Execution time	Second
Energy	Joule
Total energy	Joule
Dynamic energy	Joule

References

1. IEA. Climate Change. Available online: <https://www.iea.org/topics/climate-change> (accessed on 23 June 2020).
2. IEA. Global Energy & CO₂ Status Report 2019. Available online: <https://www.iea.org/reports/global-energy-and-co2-status-report-2019> (accessed on 17 June 2020).
3. IEA. IEA Sets Out Pillars for Success at COP21. Available online: <https://www.iea.org/news/iea-sets-out-pillars-for-success-at-cop21> (accessed on 17 June 2020).
4. Andrae, A.; Edler, T. On Global Electricity Usage of Communication Technology: Trends to 2030. *Challenges* **2015**, *6*, 117–157. [CrossRef]
5. O'brien, K.; Pietri, I.; Reddy, R.; Lastovetsky, A.; Sakellariou, R. A Survey of Power and Energy Predictive Models in HPC Systems and Applications. *ACM Comput. Surv.* **2017**, *50*, 37:1–37:38. [CrossRef]
6. Fagas, G.; Gallagher, J.P.; Gammaitoni, L.; Paul, D.J. Energy Challenges for ICT. In *ICT - Energy Concepts for Energy Efficiency and Sustainability*; Fagas, G., Gammaitoni, L., Gallagher, J.P., Paul, D.J., Eds.; IntechOpen: Rijeka, Croatia, 2017; Chapter 1.
7. Tang, Z.; Qi, L.; Cheng, Z.; Li, K.; Khan, S.U.; Li, K. An Energy-Efficient Task Scheduling Algorithm in DVFS-enabled Cloud Environment. *J. Grid Comput.* **2016**, *14*, 55–74. [CrossRef]

8. Cao, T.; He, Y.; Kondo, M. Demand-Aware Power Management for Power-Constrained HPC Systems. In Proceedings of the 2016 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), Cartagena, Colombia, 16–19 May 2016; pp. 21–31.
9. Arroba, P.; Moya, J.M.; Ayala, J.L.; Buyya, R. Dynamic Voltage and Frequency Scaling-Aware Dynamic Consolidation of Virtual Machines for Energy Efficient Cloud Data Centers. *Concurr. Comput. Pract. Exp.* **2017**, *29*, e4067. [[CrossRef](#)]
10. Lang, J.; Runger, G. An Execution Time and Energy Model for an Energy-Aware Execution of a Conjugate Gradient Method with CPU/GPU Collaboration. *J. Parallel Distrib. Comput.* **2014**, *74*, 2884–2897. [[CrossRef](#)]
11. Lastovetsky, A.; Manumachu, R.R. New Model-Based Methods and Algorithms for Performance and Energy Optimization of Data Parallel Applications on Homogeneous Multicore Clusters. *IEEE Trans. Parallel Distrib. Syst.* **2017**, *28*, 1119–1133. [[CrossRef](#)]
12. Khaleghzadeh, H.; Fahad, M.; Manumachu, R.R.; Lastovetsky, A. A Novel Data Partitioning Algorithm for Dynamic Energy Optimization on Heterogeneous High-Performance Computing Platforms. *Concurr. Comput. Pract. Exp.* **2020**, e5928. [[CrossRef](#)]
13. Fahad, M.; Shahid, A.; Manumachu, R.R.; Lastovetsky, A. Accurate Energy Modelling of Hybrid Parallel Applications on Modern Heterogeneous Computing Platforms Using System-Level Measurements. *IEEE Access* **2020**, *8*, 93793–93829. [[CrossRef](#)]
14. Fahad, M.; Shahid, A.; Manumachu, R.R.; Lastovetsky, A. A Comparative Study of Methods for Measurement of Energy of Computing. *Energies* **2019**, *12*, 2204. [[CrossRef](#)]
15. Kuzma, J.W. *Basic Statistics for the Health Sciences*, 3rd ed.; Mayfield Publishing Company: Mountain View, CA, USA, 1998.
16. Abramowitz, M. *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*; Dover Publications, Inc.: Mineola, NY, USA, 1974.
17. Dargie, W. A Stochastic Model for Estimating the Power Consumption of a Processor. *IEEE Trans. Comput.* **2015**, *64*. [[CrossRef](#)]
18. Zhou, Z.; Abawajy, J.H.; Li, F.; Hu, Z.; Chowdhury, M.U.; Alelaiwi, A.; Li, K. Fine-Grained Energy Consumption Model of Servers Based on Task Characteristics in Cloud Data Center. *IEEE Access* **2018**, *6*, 27080–27090. [[CrossRef](#)]
19. McCullough, J.C.; Agarwal, Y.; Chandrashekar, J.; Kuppaswamy, S.; Snoeren, A.C.; Gupta, R.K. Evaluating the Effectiveness of Model-Based Power Characterization. In Proceedings of the 2011 USENIX Conference on USENIX Annual Technical Conference, Portland, OR, USA, 15–17 June 2011; p. 12.
20. Hackenberg, D.; Ilsche, T.; Schone, R.; Molka, D.; Schmidt, M.; Nagel, W.E. Power Measurement Techniques on Standard Compute Nodes: A Quantitative Comparison. In Proceedings of the 2013 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), Austin, TX, USA, 21–23 April 2013; pp. 194–204.
21. Subramaniam, B.; Feng, W.C. Towards Energy-Proportional Computing for Enterprise-Class Server Workloads. In Proceedings of the 4th ACM/SPEC International Conference on Performance Engineering, Prague, Czech Republic, 21–24 April 2013; p. 1526.
22. Hackenberg, D.; Schone, R.; Ilsche, T.; Molka, D.; Schuchart, J.; Geyer, R. An Energy Efficiency Feature Survey of the Intel Haswell Processor. In Proceedings of the 2015 IEEE International Parallel and Distributed Processing Symposium Workshop, Hyderabad, India, 25–29 May 2015; pp. 896–904.
23. Khan, K.N.; Hirki, M.; Niemi, T.; Nurminen, J.K.; Ou, Z. RAPL in Action: Experiences in Using RAPL for Power Measurements. *ACM Trans. Model. Perform. Eval. Comput. Syst.* **2018**, *3*, 9:1–9:26. [[CrossRef](#)]
24. Rico-Gallego, J.A.; Dıaz-Martın, J.C.; Lastovetsky, A.L. Extending t-Lop to Model Concurrent MPI Communications in Multicore Clusters. *Future Gener. Comput. Syst.* **2016**, *61*, 6682. [[CrossRef](#)]
25. Li, B.; Han, L. *Distance Weighted Cosine Similarity Measure for Text Classification. Intelligent Data Engineering and Automated Learning – IDEAL 2013*; Yin, H., Tang, K., Gao, Y., Klawonn, F., Lee, M., Weise, T., Li, B., Yao, X., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 611–618.
26. Berndt, D.J.; Clifford, J. Using Dynamic Time Warping to Find Patterns in Time Series. In Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, 31 July–1 August 1994; pp. 359–370.
27. Nakamura, T.; Taki, K.; Nomiya, H.; Seki, K.; Uehara, K. A Shape-Based Similarity Measure for Time Series Data with Ensemble Learning. *Pattern Anal. Appl.* **2013**, *16*, 535–548. [[CrossRef](#)]

28. Box, G.E.P.; Jenkins, G. *Time Series Analysis, Forecasting and Control*; Holden-Day, Inc.: San Francisco, CA, USA, 1976.
29. Yu, Q.; Jibin, L.; Jiang, L. An Improved ARIMA-Based Traffic Anomaly Detection Algorithm for Wireless Sensor Networks. *Int. J. Distrib. Sens. Netw.* **2016**, *12*, 9653230. [[CrossRef](#)]
30. Piccolo, D. A Distance Measure for Classifying Arima Models. *J. Time Ser. Anal.* **1990**, *11*, 153–164. [[CrossRef](#)]
31. Ramoni, M.; Sebastiani, P.; Cohen, P. Bayesian Clustering by Dynamics. *Mach. Learn.* **2002**, *47*, 91–121. [[CrossRef](#)]
32. Maharaj, E.A. Cluster of Time Series. *J. Classif.* **2000**, *17*, 297–314. [[CrossRef](#)]
33. Anton, H.; Rorres, C. *Elementary Linear Algebra: Applications Version*, 11th ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2013.
34. Warren Liao, T. Clustering of Time Series Data-A Survey. *Pattern Recogn.* **2005**, *38*, 18571874. [[CrossRef](#)]
35. Iglesias, F.; Kastner, W. Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns. *Energies* **2013**, *6*, 579–597. [[CrossRef](#)]
36. Sanfeliu, A.; Fu, K. A Distance Measure between Attributed Relational Graphs for Pattern Recognition. *IEEE Trans. Syst. Man Cybern.* **1983**, *SMC-13*, 353–362. [[CrossRef](#)]
37. Fahad, M.; Shahid, A.; Manumachu, R.R.; Lastovetsky, A. Supplementary Materials: A Novel Statistical Learning-Based Methodology for Measuring the Goodness of Energy Profiles of Applications Executing on Multicore Comptuing Platforms. Available online: https://csgitlab.ucd.ie/Muhammad_Fahad/supplementals-to-the-publications/-/blob/master/2020/MDPI_goodness/supplemental.pdf (accessed on 25 July 2020).
38. Rotem, E.; Naveh, A.; Ananthakrishnan, A.; Weissmann, E.; Rajwan, D. Power-Management Architecture of the Intel Microarchitecture Code-Named Sandy Bridge. *IEEE Micro* **2012**, *32*, 20–27. [[CrossRef](#)]
39. Heterogeneous Computing Laboratory, University College Dublin. HCLWattsUp: API for Power and Energy Measurements Using WattsUp Pro Meter. Available online: <https://csgitlab.ucd.ie/ucd-hcl/hclwattsup> (accessed on 23 May 2020).
40. Rawlings, J.O.; Pantula, S.G.; Dickey, D.A. *Applied Regression Analysis*, 2nd ed.; Springer: New York, NY, USA, 1998.
41. Ostertagová, E. Modelling Using Polynomial Regression. *Procedia Eng.* **2012**, *48*, 500–506. [[CrossRef](#)]
42. Shahid, A.; Fahad, M.; Manumachu, R.R.; Lastovetsky, A. Energy of Computing on Multicore CPUs: Predictive Models and Energy Conservation Law. Available online: <https://arxiv.org/pdf/1907.02805.pdf> (accessed on 15 June 2020).
43. Nvidia. NVML Reference Manual. Available online: https://docs.nvidia.com/pdf/NVML_API_Reference_Guide.pdf (accessed on 16 June 2020).
44. Corporation, I. Intel Xeon Phi Coprocessor SystemSoftware Developers Guide. Available online: <https://software.intel.com/sites/default/files/managed/09/07/xeon-phi-coprocessor-system-software-developers-guide.pdf> (accessed on 16 June 2020).
45. Corporation, I. Intelo Manycore Platform Software Stack (Intel MPSS). Available online: <https://software.intel.com/en-us/articles/intel-manycore-platform-software-stack-mpss> (accessed on 16 June 2020).



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).