

Composite Measures of Supercomputer Technology

Nam-Gyu Kim¹, Noo Ri On¹, Myoung-Ju Koh¹, JongSuk Ruth Lee^{1*} and Keun-Tae Cho^{2*}

¹Division of National Supercomputing, Korea Institute of Science and Technology Information
Daejeon - ROK

[e-mail: ssgyu, claireon, myju, jsruthlee@kisti.re.kr]

²Department of Management of Technology, Sungkyunkwan University
Suwon-Si, Gyeonggi-Do - ROK
[e-mail: ktcho@skku.edu]

*Corresponding author: JongSuk Ruth Lee and Keun-Tae Cho

*Received March 3, 2019; revised May 25, 2019; accepted June 6, 2019;
published August 31, 2019*

Abstract

We have developed composite measures of supercomputer technology, reflecting various factors of supercomputers using Martino's scoring model. CPUs, accelerators, memory, interconnection networks, and power consumption are chosen as factors of the model. The weight values of the factors are derived based on a survey of 129 domestic and international experts. The measured values are then standardized to integrate measurement units of the factors in the model. This model has been applied to 50 supercomputers, and rank correlation analysis was performed using representative measures. As a consequence, the ranking drastically changes except for the 1st and 2nd supercomputers on the TOP500. In addition, the characteristics of memory and interconnection networks influence the ranking, and the results demonstrate that the proposed model has low correlations with HPL and HPCG but a high correlation with Green500. This indicates that power consumption is an important factor that has a significant effect on the measures of supercomputer technology. In addition, it is determined that the differences between the HPL ranking and the proposed model ranking are influenced by power consumption, CPU theoretical peak performance, and main memory bandwidth in order of significance. In conclusion, the composite measures proposed in this study are more suitable for comprehensively describing supercomputer technology than existing performance measures. The findings of this study are expected to support decision making related to management and policy in the procurement and operation of supercomputers.

Keywords: Composite Measures, Measures of Technology, Scoring Model, Supercomputer Performance, Supercomputer policy

A preliminary version of this paper was presented at ICONI 2018, and was selected as an outstanding paper. This research was supported by the KISTI Program (No. K-19-L02-C05-S01), the EDISON Program through the National Research Foundation of Korea (NRF) (No. NRF-2011-0020576), the NRF and the Center for Women In Science, Engineering and Technology (WISSET-2019-233). A Grant was awarded by the Ministry of Science and ICT (MSIT) under the Program for Returners for R&D.

1. Introduction

Scientific computing or computational science and engineering is a multidisciplinary field comprising algorithms, software, computers, information technology, and computing infrastructure required to understand and solve complex problems in science and engineering [1,2,3]. Supercomputers are computer systems capable of the most powerful computation performance for solving such problems. Therefore, the performance of supercomputers is an important concern for researchers and organizations that own and operate such computer systems [3,4].

The TOP500 Project, which began in 1993, ranks the top 500 supercomputers in the world and announces the ranking twice a year based on the results of the benchmark program High-Performance LINPACK (HPL) [5,6]. HPL is easy to measure, use, and update because it indicates the performance of supercomputers based on the number of floating-point operations per second (FLOPS). However, accelerators, such as GPU and many-core processors, are currently being employed for data-based artificial intelligence studies such as deep learning and big data analysis studies that require massive data processing [7]. In the life sciences field, memory capacity and memory bandwidth limitations present considerable challenges as a significant amount of memory is needed for large-scale genetic analysis [7,8]. Consequently, HPL has been strongly criticized for low correlations with the operating performance of applications because the measurements of HPL do not reflect the characteristics of other important factors of supercomputers [9-12].

The High Performance Conjugate Gradients (HPCG) benchmark that complements HPL reflects the characteristics of data access patterns of current applications. However, HPCG exhibits a shortcoming in which benchmark results are greatly affected by actual memory bandwidth [11,12]. In addition, other supercomputer benchmarking projects for specific purposes are in progress, such as Graph500, which focuses on supercomputer performance measurements for data-intensive applications, and Deep500, which is intended for performance measurement of deep learning applications.

Meanwhile, as supercomputer performance increases, power consumption also increases, raising concerns associated with “The Energy Crisis in Supercomputing” [13,14]. Therefore, power consumption is a critical factor that must be considered in the procurement and operation of supercomputers, and as a consequence, accelerators are increasingly being used in many supercomputers because of their power efficiency and high parallel computing performance characteristics [15-20]. The Green500, which measures the energy efficiency of supercomputers based on FLOPS per watt, has been launched through the TOP500 Project [21]. This means that Green500 denotes the HPL measurement of per unit power, therefore, it also contains the disadvantages of HPL.

This study proposes composite measures of supercomputer technology using a scoring method that indicates the technology level of supercomputers by comprehensively considering the utilization characteristics of users and the primary factors of supercomputers. This paper suggests that the model proposed in this study is suitable for comprehensively describing the technology level of supercomputers compared to existing simple performance benchmark program-based indicators, and that this can help to inform management decisions and policies for procurement and operation of supercomputers.

The remainder of this study is organized as follows. Section 2 presents a discussion on prior studies on representative performance measurement methods for supercomputers including

their limitations. Section 3 presents the proposed research methods and composite measures of supercomputer technology. Section 4 evaluates supercomputers based on the proposed composite measures and compares the ranking order of the supercomputers derived from the proposed model with the existing ranking orders of supercomputers using traditional benchmarks. Section 5 summarizes the research results and presents a discussion on the limitations and future works.

2. Prior Studies on Supercomputer Performance Measurement

The performance of supercomputers is a critical concern for researchers who use them for solving complex and large problems in science and engineering. Various benchmarks have been developed to measure the performance of supercomputers, but the most widely used benchmarks are HPL and HPCG.

HPL is a benchmark program that determines the solution to $Ax = b$, which denotes a large-scale dense matrix problem of a linear equation. The performance of HPL is determined by the 64-bit floating-point operation used in multiplication of the dense matrix, which is a major calculation in the methodology of the benchmark program [5,22]. The FLOPS value, which is obtained from HPL, is used as a measure of supercomputing performance in the TOP500 Project, which presents a list of the top 500 fastest supercomputers in the world since 1993. When HPL was initially applied in the TOP500 Project, most benchmark applications were based on the dense matrix calculation, which is similar to HPL. However, the majority of current applications compute differential equations that require high memory bandwidth and irregular data access. As a consequence, there is a low correlation between the performances of HPL and the application [22,23,24].

To complement this disadvantage of HPL, HPCG was developed by Jack Dongarra and the researchers who developed HPL. HPCG is a benchmark program that solves a partial-differential equation in a three-dimensional grid discretized with a 27-point stencil. Similar to HPL, HPCG determines the solution to $Ax = b$, but unlike HPL, the matrix A is sparse. The solution process uses the conjugate gradient algorithm and the Gauss–Seidel preconditioner, unlike HPL [23]. HPCG compensates for the shortcoming of the HPL by reflecting the characteristics of data access patterns seen in current applications. The result of HPCG depends on the actual bandwidth of memory because HPCG requires an excessively larger memory bandwidth compared to real applications, which is a shortcoming of HPCG [13,14].

Compared to supercomputer performance evaluation obtained from benchmark programs, which comprise approaches based on the concept that performance is denoted by calculation speed, the Green500 Project focuses on the power efficiency of supercomputers and derived the ranking based on FLOPS per watt. Because the increase in calculation performance of supercomputers has caused a rapid increase in power consumption, power efficiency presents an important factor that must be considered when procuring and operating supercomputers, leading to the concept of “Green Supercomputing” [13,14]. For example, the Earth Simulator in Japan, which ranked no.1 for two years from 2002, incurred an annual cost of 10 million US dollars (USD) to run at 12 MW power consumption [13]. Moreover, Nurion, the 5th generation supercomputer introduced at KISTI in Korea in 2018, incurs a peak power consumption of 3809 kW, and the annual electricity cost incurred by Nurion is estimated to be 4.4 billion KRW. In general, 1MW-year worth of electricity costs approximately one million USD, and power consumption costs are increasing uncontrollably as the performance of supercomputers increases, owing to competition among developed countries [13,14,19,20]. However,

Green500 is also a straightforward metric that presents the computation performance per unit power by dividing the HPL benchmark results of supercomputers by their power consumption. Therefore, Green500 also includes the shortcomings of HPL.

3. Research Method and Model

The HPL, HPCG, and Green500 in the TOP500 project are largely influenced by specific components of supercomputers and hold the disadvantage of not reflecting the features of various supercomputer components. Supercomputers consist of various factors, and supercomputer technology depends on the combination of these factors. In addition, there is no specific analysis method that combines factors for composite measures of supercomputer technology. Therefore, developing comprehensive composite measures that overcome the limitations of previous studies is essential. In this paper, we use Martino's general scoring model, which has been widely used to develop composite measures based on a combination of several factors [25,26].

To develop a scoring model, factors must be first identified. All factors that are necessary for the technology to function properly are listed. Duplicate factors are excluded and only measurable factors are chosen. Similar factors are then grouped together. Weightings are then assigned to each factor by considering the relative importance of the grouped factors. Standardization is then performed to convert all units to the same scale, which can be used to evaluate and measure all factors.

3.1 Selection and Identification of Factors

In this paper, we focus on supercomputer hardware to measure the technology level of the supercomputer regardless of its specific applications. Therefore, we select several major hardware components of the supercomputer as factors for the scoring model. Major hardware components of the supercomputer such as processor, memory, interconnection networks, I/O system, and storage are considered based on several previous studies [4,12,27,28]. As described above, the performance of a factor must be objectively measurable. Because most supercomputer manufacturers do not publish information on the internal I/O systems in the computer node, the I/O systems were excluded. Storage systems were also excluded because it is difficult to measure the storage capacity used by a specific HPC system, because several supercomputers use several storages in most organizations. In conclusion, we selected processors, memory, and interconnection networks as the major factors, which can provide objective data. Data on them are available from the TOP500 project, which includes history and credibility as supercomputer statistical data repository dating back to 1993. Furthermore, to reflect recent trends, processors are divided into CPUs and accelerators such as graphics processing units (GPUs) and a many integrated core (MIC) architecture.

In addition, as described in the previous section, the power consumption of supercomputers has been added as a major factor because it is an important consideration for supercomputer procurement and operation. Information on accelerators and power consumption are also included in the TOP500 project.

3.2 Weights of the Factors

To derive the weights assigned to the major factors presented in this paper, we conducted a survey on the importance of major supercomputer components to domestic and foreign researchers. In total, 129 questionnaires were collected during supercomputing conferences from computational science and engineering researchers who use high-performance

computers (cluster, cloud, or supercomputers). **Fig. 1** presents the descriptive statistics of the survey respondents.

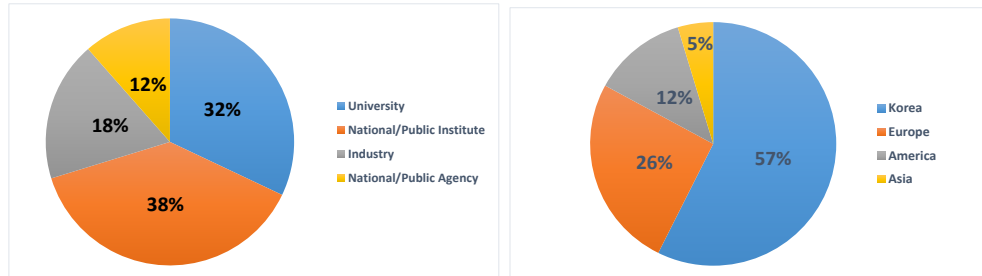


Fig. 1. Respondents' affiliations and regions

Based on the survey, we performed multiple ranked response analysis to determine the importance of CPUs, accelerators, memory, and interconnection networks. The frequencies of the 1st, 2nd, 3rd, and 4th rankings based on the importance of each factor are calculated, and the significance is assigned to each ranking multiplied by 4, 3, 2, or 1. The proportions of sums are used as weights [29]. **Table 1** summarizes the results.

Table 1. Rankings and weights based on importance of major supercomputer components

Component	1st	2nd	3rd	4th	Sum of Importance	%	Weight	Symbol
CPU	80	30	14	5	443	34.34	0.35	ω_p
Accelerator	21	22	40	46	276	21.40	0.21	ω_a
Memory	22	53	39	15	340	26.36	0.26	ω_m
Interconnection networks	6	24	36	63	231	17.91	0.18	ω_i

In addition, we consider the performance of the supercomputer in terms of its effectiveness while considering power consumption in terms of costs. We then derive the weights for performance and power consumption. The procurement costs and electricity costs of the top 10 supercomputers among 50 target supercomputers were calculated. Procurement costs were sourced through the Internet, and the electricity cost is estimated at 1 million USD per 1MW-year [11-14]. As described above, because supercomputer performances are increasing rapidly, the lifetime of supercomputers is considered 5 years, so the electricity cost has been calculated for 5 years. [4,7]. Therefore, the ratio of the cost of procurement and operation is determined by assuming that the sum of the procurement price and the electricity cost for 5 years is the total cost of each supercomputer. Based on the average across the 10 supercomputers, we determined the ratios of procurement cost and operating cost to total cost to be 0.81 and 0.19, respectively. Therefore, the system performance weight(ω_e) is 0.81 and the power consumption weight(ω_c) is 0.19. **Table 2** summarizes the results.

Table 2. Procurement costs and electricity costs of the top 10 supercomputers

TOP500 Ranking	Supercomputers	Power Consumption (kw)	Procurement Cost (million USD)	5-year electricity cost (million USD)	Procurement Cost Ratio	Operating Cost Ratio
1	Summit	8,806	200	44.03	0.819571	0.180429
2	Sunway TaihuLight	15,371	273	76.855	0.780323	0.219677

7	Titan	8,209	97	41.045	0.702669	0.297331
8	Sequoia	7,890	250	39.45	0.863707	0.136293
9	Trinity	3,844	174	19.22	0.900528	0.099472
10	Cori	3,939	70	19.695	0.780423	0.219577
11	Nurion	3,809	48	19.045	0.715937	0.284063
16	K Computer	12,660	1,000	63.3	0.940468	0.059532
17	Mira	3,945	50	19.725	0.717103	0.282897
29	Shaheen II	2,834	80	14.17	0.849527	0.150473
Sum			2242	356.535	8.070257	1.9297427
Ratio			0.862794	0.137206	0.81	0.19

3.3 Construction of the Model

The composite measure model for reflecting major supercomputer factors is based on the general scoring model developed by Joseph Martino. The model proposed in this study is expressed as follows:

$$\frac{(\alpha P + \beta A)^a (\lambda MS \cdot MB + \delta AS \cdot AB)^b I^c}{PW^d} \quad (1)$$

Here, $\alpha + \beta = 1, a + b + c + d = 1$

In the equation (1), uppercase letters denote factors, and lowercase letters denote the weights applied to the factors. The uppercase letters of the numerator indicate the desirable factors, and the larger the values of factors, the higher the score of the technology. The uppercase letters of the denominator denote costs or other undesirable factors. If these values increase, the score of the technology decreases. **Table 3** presents a description of the factors in the model.

Table 3. Factors Considered in the Model

Symbol	Definition	Unit of Measurement
P	CPU theoretical peak performance	GFLOPS
A	Accelerator theoretical peak performance	GFLOPS
MS	Main memory size	GB
MB	Main memory bandwidth	GB/s
AS	Accelerator memory size	GB
AB	Accelerator memory bandwidth	GB/s
I	Interconnection network bandwidth	GB/s
PW	Power consumption	kW

P, A, MS, MB, AS, AB, and I are placed in the numerator because the performance of supercomputers increases as their values increase. PW is placed in the denominator because it denotes the cost of these supercomputers.

I is an overriding factor, which is the most important and is expressed as a multiplication in the equation. The total score is computed as zero when any of these factors is zero.

P and A, which are related to computation performance, and MS, MB, AS, and AB, which are related to memory performance, are tradable factors. Even if the value of any of these factors is zero, the total score is not zero. These factors can be expressed as a multiplication in the equation after integration with other factors.

The weight values a , b , c , d , α , β , γ , and δ are determined by expert judgment or statistics. Their sum must be 1 because the weight values of the factors need to be normalized. This study also considers a model that did not consider power consumption for comparing the rankings of the conventional HPL and HPCG benchmarks. Consequently, the weights are different between the models that consider power consumption and those that do not. The method of determining the weight values of factors a , b , c , and d is described in [Table 4](#).

Table 4. Weights of the Model Factors

Symbol	Definition	Power consumption included		Power consumption excluded	
		Equation	Weight	Equation	Weight
a	The weight of computation performance	$(\omega_p + \omega_a)^{\omega_e}$	0.45	$\omega_p + \omega_a$	0.56
b	The weight of memory performance	$(\omega_m)^{\omega_e}$	0.21	ω_m	0.26
c	The weight of interconnection network performance	$(\omega_i)^{\omega_e}$	0.15	ω_i	0.18
d	The weight of power consumption	ω_c	0.19		

α , β , γ , and δ are the weight values of tradable factors. The sum of the theoretical peak performance of CPUs and accelerators is the theoretical peak performance of a supercomputer. Therefore, P and A are expressed by an addition, and their weight values α and β are denoted by the proportions of the total theoretical performance. Furthermore, the sum of the main memory performance and the accelerator memory performance is the total memory performance of the supercomputer. γ and δ indicate the weight values that denote their importance, and they are equal to α and β because they are determined by the proportions of the CPU and accelerator that make up the total theoretical performance in the supercomputer. The equations for calculating the weights of the tradable factors are listed in [Table 5](#).

Table 5. Weights between the tradable factors

Symbol	Definition	Equation	Weight
α	The CPU's proportion of computation performance	$\frac{\omega_p}{\omega_p + \omega_a}$	0.62
β	The accelerator's proportion of computation performance	$\frac{\omega_a}{\omega_p + \omega_a}$	0.38

γ	The main memory's proportion of memory performance	$\frac{\omega_p}{\omega_p + \omega_a}$	0.62
δ	The accelerator memory's proportion of memory performance	$\frac{\omega_a}{\omega_p + \omega_a}$	0.38

Finally, we propose the power model (P-Model) by integrating the factors and their weights. On the other hand, to verify the validity of the P-Model, we also proposed the non-power model (NP-Model) that does not include the power consumption factor for comparing against representative supercomputer performance measures. The models are defined as follows:

$$\frac{(0.62P + 0.38A)^{0.45} (0.62MS \cdot MB + 0.38AS \cdot AB)^{0.21} I^{0.15}}{PW^{0.19}} \quad (\text{P-Model})$$

$$(0.62P + 0.38A)^{0.56} (0.62MS \cdot MB + 0.38AS \cdot AB)^{0.26} I^{0.18} \quad (\text{NP-Model})$$

3.4 Centering and Standardizing the Measured Values

Each factor uses a different measurement unit or measures a different value depending on the measurement unit, even for the same factor. For example, the measurement unit of the CPU is FLOPS, which is different from B/s as the measurement unit of interconnection networks. In addition, the value can differ by 1,000 times depending on whether GFLOPS or TFLOPS are used as the measurement unit of CPU peak performance. Therefore, the measurements of factors with different measurement units and values must be standardized to construct a composite measure. The mean and normal distribution of the measured values must be calculated for each factor, and the measured values of each factor must be standardized based on them. Finally, to center the standardized values onto a scale within the same range, the standardized values were multiplied by 0.7 and then 5 added so that the values fall between 1 and 10.

4. Analysis and Results

4.1 Target Selection

To apply the scoring model proposed in this paper, we selected supercomputers with measured values of HPL, HPCG, and Green 500 on the TOP500 list as of June 2018 to compare our proposed model with representative supercomputer performance measures. Moreover, Nurion, the latest supercomputer to be introduced at KISTI National Supercomputing Division in Korea, was added to the analysis target. Fifty supercomputers, excluding machines with no values for the major factors among the selected supercomputers, were finally selected for analysis.

4.2 Analysis Results

The P-Model was used as the composite measure model of supercomputer technology, and the NP-Model was used for comparisons with representative measures of the selected 50 supercomputers. Table 6 presents the rankings of the 50 supercomputers that were selected based on the P-Model score and the NP-Model score, including their HPL, HPCG, and Green 500 rankings and the differences between the HPL ranking and the P-Model ranking.

Summit and Sunway TaihuLight, which ranked the 1st and 2nd in the P-Model ranking, were also ranked the 1st and 2nd in the TOP 500, respectively. These are overwhelmingly larger in scale than other supercomputers because Summit and Sunway TaihuLight have maximal achieved performance of 122.3 PF and 93 PF on HPL, respectively [6]. Therefore, their scores were much higher than those of other supercomputers. Trinity, which ranked 6th in the HPL ranking, was 3rd in the P-Model ranking. Trinity contains 9,984 nodes, which are equipped with an Intel Xeon Phi Knight Landing with high-bandwidth memory (MCDRAM) with 16 GB capacity and a peak memory bandwidth of over 450 GB/s, nominally five times faster than DDR4. Consequently, Trinity has a large memory capacity of over 2 PB [29]. Trinity was the third-ranked supercomputer in the P-Model because it consists of top-ranked factors such as main memory size, accelerator peak performance, and accelerator memory size, which were ranked 2nd among the 50 supercomputers. Plasma Simulator, which ranked 4th in the P-Model, consists of a Fujitsu PRIMEHPC FX 100 with HMC (Hybrid Memory Cube), which has a high memory bandwidth of 240 GB/s x in/out [6,31]. In addition, this has the Tofu 2 interconnection network with a high bandwidth of 50 GB/s per node [32]. The supercomputer at the Information Technology Center, Nagoya University, which ranked 5th, and the SORA-MA, which ranked 6th in the P-Model ranking, also consist of the same hardware features as Plasma Simulator [33]. Based on the results, the P-Model 4th, 5th, and 6th ranked systems, which have risen significantly from their HPL rankings, have the characteristics of systems in which the main memory bandwidth and the interconnection network bandwidth are ranked 1st among the 50 supercomputers, respectively.

In other words, the result demonstrates that the model presented in this paper reflects not only the performance of CPUs and accelerators but also the memory bandwidth that is important in the current application, and the interconnection network bandwidth for data transfer between nodes that are important in massively parallel applications.

In the case of Nurion, the HPL ranking and the HPCG ranking were 8th and 6th in the analysis target 50s, respectively. By contrast, it was ranked 16th on the Green500 ranking, and so the efficiency of power consumption by Nurion was lower than that of other supercomputers. However, Nurion ranked 12th in the P-Model as the performance and power consumption of Nurion were properly reflected.

Table 6. Scores and rankings of the proposed models and representative supercomputer rankings

Supercomputer	HPL Ranking	HPCG Ranking	Green500 Ranking	P-Model Ranking	NP-Model Ranking	The differences between the HPL ranking and the P-Model ranking	P-Model Score	NP-Model Score
Summit	1	1	1	1	1	0	3.4581	10.7886
Sunway TaihuLight	2	5	8	2	2	0	3.1418	10.2427
Piz Daint	3	4	3	15	16	-11	2.9112	7.7394
Titan	4	11	30	44	17	-28	2.7116	7.6785
Sequoia	5	10	27	39	14	-17	2.7650	7.8584
Trinity	6	3	15	3	3	3	3.0790	8.5694
Cori	7	8	17	14	8	-5	2.9312	8.0250
Nurion	8	6	16	12	10	-5	2.9322	8.0129
Oakforest-PACS	9	7	9	11	11	-2	2.9636	7.9871

HPC4	10	12	5	10	13	0	2.9852	7.8761
Tera-1000-2	11	9	14	9	4	3	3.0046	8.1980
K Computer	12	2	47	50	15	-33	2.6271	7.7739
Mira	13	15	28	33	21	-11	2.7782	7.4674
TSUBAME3.0	14	13	2	7	9	7	3.0435	8.0163
MareNostrum	15	18	13	19	18	-1	2.8830	7.5552
Pleiades	16	14	44	43	27	-23	2.7262	7.3267
Hazel Hen	17	17	42	30	22	-6	2.7828	7.4417
Shaheen II	18	19	36	24	25	-7	2.7912	7.3782
Pangea	19	16	45	48	38	-30	2.6670	7.0843
Cheyenne	20	21	20	28	32	-11	2.7853	7.2212
Vulcan	21	24	29	29	29	-9	2.7837	7.2480
JOLIOT-CURIE SKL	22	33	10	18	20	5	2.8873	7.4741
ECMWF1	23	30	32	25	30	-4	2.7870	7.2475
ECMWF2	24	31	33	25	30	-3	2.7870	7.2475
Electra	25	32	18	20	23	5	2.8672	7.4105
DGX Saturn V	26	29	6	8	12	17	3.0258	7.8879
SORA-MA	27	20	37	6	5	21	3.0592	8.1954
Mistral	28	36	21	38	40	-14	2.7656	7.0741
ICT-Nagoya	29	22	31	5	6	25	3.0670	8.1849
SuperMUC	30	23	46	49	48	-20	2.6655	6.9962
TSUBAME 2.5	31	28	35	42	42	-12	2.7500	7.0555
occigen2	32	35	41	40	37	-4	2.7637	7.1066
Plasma Simulator	33	27	38	4	7	28	3.0674	8.1663
Prolix2	34	40	22	36	44	-6	2.7702	7.0531
Beaufix2	35	39	23	35	43	-3	2.7707	7.0549
Tera-1000-1	36	41	40	37	39	-1	2.7687	7.0749
Prometheus	37	37	34	22	33	11	2.8066	7.1756
Edison	38	26	50	47	35	-9	2.6891	7.1180
ARCHER	39	25	49	45	36	-7	2.7014	7.1121
Sid	40	49	24	34	46	-1	2.7734	7.0272
Curie thin nodes	41	34	48	46	50	-7	2.7004	6.9701
JOLIOT-CURIE KNL	42	38	12	16	24	24	2.8907	7.4034
Cobalt	43	43	25	23	34	14	2.8048	7.1358
Graham	44	46	26	21	28	23	2.8585	7.3220
Cartesius2	45	48	43	41	49	1	2.7574	6.9931
Falcon	46	47	19	32	45	11	2.7810	7.0285
EAGLE	47	50	39	31	41	13	2.7818	7.0568
Reedbush-L	48	42	4	13	19	33	2.9312	7.5127
Reedbush-H	49	45	7	17	26	30	2.8880	7.3625
XStream	50	44	11	27	47	17	2.7868	7.0268

4.3 Comparisons with Representative Supercomputer Measures

To compare the proposed composite measures with existing measures of supercomputers, a rank correlation analysis was performed on the rankings from [Table 6](#) and the results are listed in [Table 7](#).

Table 7. Results of rank correlation analysis

Symbol	Definition	HPL Ranking	HPCG Ranking	Green500 Ranking	P-Model Ranking	NP-Model Ranking
HPL Ranking	Correlation coefficient	1.000				
	Probability of significance		0.000			
HPCG Ranking	Correlation coefficient	.930**	1.000			
	Probability of significance	0.000				
Green 500 Ranking	Correlation coefficient	0.267	0.157	1.000		
	Probability of significance	0.061	0.275			
P-Model Ranking	Correlation coefficient	.290*	0.271	.717*	1.000	
	Probability of significance	0.041	0.057	0.000		
NP-Model Ranking	Correlation coefficient	.692*	.725**	.476**	.759**	1.000
	Probability of significance	0.000	0.000	0.000	0.000	

** Spearman correlation is significant at the 0.01 level (2-tailed)

* Spearman correlation is significant at the 0.05 level (2-tailed)

We first examined the ranking of conventional measures of supercomputers. The correlation between the HPL ranking and the HPCG ranking showed a correlation coefficient of 0.93 and a significance level of less than 0.01, indicating a very high correlation [34]. Therefore, a supercomputer with a high HPL ranking is expected to have a high HPCG ranking as well, and the explanatory power is 86%. However, the correlation of the HPL ranking and the Green500 ranking showed a correlation coefficient of 0.271 and a significance level of 0.061, indicating low correlation. In addition, the correlation coefficient of the HPCG ranking and the Green500 ranking was not significant. This means that the ranking of a supercomputer can change considerably when power consumption is included in the measures, and the measures that include power consumption have a different explanatory power compared to the simple performance benchmarks.

The correlation between the P-Model ranking and the HPL ranking showed a correlation coefficient of 0.29 and a significance level of 0.41. In addition, the correlation of the P-Model ranking and the HPCG ranking showed a correlation coefficient of 0.271 and a significance level of 0.057. This means that the P-Model ranking is very different from the rankings derived by the HPL and HPCG benchmark results. In contrast, there was a high correlation between the P-Model ranking and the Green500 ranking, and the explanatory power between them was 51.4% as the correlation coefficient between them was 0.717 and the significance level was less than 0.01.

The ranking of the NP-Model for comparing with existing measures showed correlations with the HPL ranking and the HPCG ranking, with correlation coefficients of 0.692 and 0.725, respectively and a significance level of less than 0.01. However, there was a low correlation

between the rankings of the NP-Model and Green 500, which showed a correlation coefficient of 0.29 and a significance level of less than 0.01.

In summary, there was no correlation of the HPL ranking and the HPCG ranking with the Green500 ranking because power consumption was included only in the latter measure. Likewise, the correlations of the HPL ranking and the HPCG ranking with the P-Model were very low because power consumption was only included in the P-Model. By contrast, the correlations of the HPL ranking and the HPCG ranking with the NP-Model were high, as none of the models considered power consumption. This suggests that the power consumption of a supercomputer has a greater effect on rankings compared with other factors. In other words, power consumption is a critical factor in the evaluation of supercomputer technology.

The above analysis demonstrates that the composite measures of supercomputers proposed in this study have high explanatory power for the overall technology level of supercomputers because they include a better variety of influencing factors compared to the simple measures used by conventional benchmarks.

4.4 Analysis on the differences between the HPL Ranking and the P-Model Ranking

To identify factors in terms of differences between the P-Model and existing HPC performance measures, the differences between the P-Model ranking and existing rankings were analyzed. For this purpose, only variation between the HPL ranking and the P-Model ranking was analyzed. In the above section, the rank correlation analysis results demonstrated that the HPL and the HPCG rankings were similar, and the ranking variation between the P-Model and the Green500 had an approximately 50% explanatory power. Therefore, it is easy to determine which factors have a significant influence on rank changes by analyzing rankings which are considerably different. In addition, it is meaningful to identify the cause of the ranking differences between the P-Model proposed in this paper and the most representative existing measure.

First, the differences between the P-Model ranking and the HPL ranking of the 50 supercomputers were used to derive the ranking change values. Then, to evaluate the overall tendency of the ranking changes, these values were divided into five levels based on the quartile with a significant drop, slight drop, no change, slight rise, or significant rise. As independent variables, we used all eight major factors employed in the P-Model. Because the ranking change level, as the dependent variable, included two or more ranks, ordinal logistic regression analysis was performed [34].

Prior to conducting the ordinal logistic regression analysis, the analysis must satisfy the basic assumption that each independent variable has the same effect when a dependent variable changes by one unit. As shown in **Table 8**, the results of the parallel lines test demonstrated that the significance level is 0.527, higher than 0.05, and so it is possible to perform ordinal logistic regression analysis.

Table 8. Results of the parallel line test

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	98.360			
General	75.474	22.886	24	0.527

The results of the model fitting test were presented in **Table 9**. The log-likelihoods of the intercept-only model and the final model with independent variables were 148.491 and 98.360, respectively. The chi-square statistic, the difference between the two models, was 50.131,

which was at a significance level of 0.00. Therefore, this model is valid because the final model, in which independent variables are added, is more significant for explaining ranking variations.

Table 9. Result of model fitting test

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	148.491			
Final	98.360	50.131	8	0.000

The goodness-of-fit test was performed to confirm that the data for analysis is well suited to a given model. The probabilities of significance for Pearson and Deviance statistics were .914 and 1.000, respectively, which are both significance levels greater than 0.05. Therefore, the data were well suited to a given model as shown in **Table 10**.

Table 10. Results of goodness-of-fit

	Chi-Square	df	Sig.
Pearson	158.468	184	0.914
Deviance	98.360	184	1.000

The pseudo R^2 , which is the coefficient of determination in a logistic regression model, generally has a lower value than the normal regression model, even if the model is well fitted. The Pseudo R^2 values for the Nagelkerke, Cox and Snell, and McFadden were 0.667, 0.633, and 0.338, respectively as shown in **Table 11**.

Table 11. Results of pseudo R-square

Class	pseudo R-square
Nagelkerke	0.667
Cox and Snell	0.633
McFadden	0.338

The results of the ordinal logistic analysis are shown in **Table 12**. Only CPU theoretical peak performance, main memory bandwidth, and power consumption were found to be significant independent variables with a significance level of less than 0.01. The larger Wald statistic means that an independent variable has a greater influence on the dependent variable. Therefore, the differences between the HPL ranking and the P-Model ranking was influenced by power consumption, CPU theoretical peak performance, and main memory bandwidth in order of significance.

CPU theoretical peak performance is a positive sign of the coefficient estimate, and if CPU theoretical peak performance increases by one unit, the odds of being included in the one-level higher category of the ranking change are increased by 0.000014249335%. This result occurred because the measured values of CPU theoretical peak performance were extreme, ranging from a minimum of 29,120 to a maximum of 128,446,365. In addition, in the ordinal logistic regression analysis, when continuous variables, as independent variables, are used as inputs, the odds ratio, which indicates changes in the dependent variable, is very low. Main memory bandwidth is a positive sign of the coefficient estimate, and if main memory bandwidth increases by one unit, the odds of being included in the one-level higher category of the ranking change are increased by 3.2%. Power consumption is a negative sign of the coefficient estimate and if power consumption increases by one unit, the odds of being included in the one-level higher category of the ranking change are reduced by 0.002%.

Table 12. Results of the ordinal logistic analysis

Variables	B estimate	S. E.	Wald	df	Sig.	95% Confidence interval		OR (Odds Ratio)
						Lower bound	Upper bound	
CPU theoretical peak performance	0.000	0.000	7.337	1	0.007	0.000	0.000	1.000
Accelerator theoretical peak performance	0.000	0.000	2.989	1	0.084	0.000	0.000	1.000
Main memory size	0.000	0.000	3.655	1	0.056	0.000	0.000	1.000
Main memory bandwidth	0.031	0.012	6.780	1	0.009	0.008	0.055	1.032
Accelerator memory size	0.000	0.000	1.243	1	0.265	0.000	0.000	1.000
Accelerator memory bandwidth	0.003	0.002	3.104	1	0.078	0.000	0.007	1.003
Interconnection network bandwidth	0.000	0.005	0.000	1	0.985	-0.009	0.010	1.000
Power consumption	-0.002	0.000	11.039	1	0.001	-0.002	-0.001	0.998

In summary, the results of the ordinal logistic analysis on the differences between the HPL ranking and the P-Model ranking demonstrated that CPU theoretical peak performance and main memory bandwidth have a positive (+) effect, but power consumption has a negative effect. In addition, power consumption has the greatest influence on ranking changes, followed by CPU theoretical peak performance and main memory bandwidth. Therefore, supercomputers with low power consumption, high main memory bandwidth, and high CPU theoretical peak performance are more likely to rank higher in the P-Model than in HPL. This is consistent with the results that the systems ranked 4th, 5th, and 6th in P-Model, higher than in the HPL ranking, had characteristics in which the main memory bandwidth was ranked first among 50 supercomputers.

5. Conclusion

In this paper, we developed the composite measures of supercomputer technology that reflected the characteristics of a variety of supercomputer components to overcome the fragmentation of traditional supercomputer performance measures. For this purpose, major supercomputer hardware components such as CPUs, accelerators, interconnection networks, and power consumption were chosen as factors. Weightings were then derived by conducting surveys of domestic and foreign experts' opinions regarding which factors were important. Furthermore, we applied the concept of cost-effectiveness to derive weights to represent the importance of supercomputer performance and power consumption. Finally, we developed the model by integrating the factors and weights based on Martino's general scoring model.

As a consequence of applying this model to selected supercomputers on the TOP500 list, the rankings significantly changed except for the supercomputers ranked 1st and 2nd. In addition,

we found that the characteristics of the memory and the interconnection networks, which were not reflected in the conventional measures, influenced the ranking. In addition, the result of a rank correlation analysis on the proposed model and representative measures' rankings demonstrated that the proposed model showed low correlations with the HPL and the HPCG ranking, but a high correlation with the Green500 ranking. This indicated that power consumption is a critical factor that significantly influences supercomputer technology. Furthermore, the differences between the HPL ranking and the P-Model ranking was found to be influenced by power consumption, CPU theoretical peak performance, and main memory bandwidth in order of significance. In conclusion, we demonstrated that the composite measures presented in this study are suitable to comprehensively measure the supercomputer technology compared to conventional measures based on the benchmarks. The proposed composite measures are expected to contribute to management decisions and policies in the procurement and operation of supercomputers in the future.

However, it is important to note that it may result in bias, depending on the selected targets, when this model is applied in practice. Supercomputers are state-of-the-art facilities whose price and performance change rapidly with time. Therefore, including older systems in the analysis may result in bias resulting from significant performance gaps. For example, when calculating the ratio of the procurement cost to the operating cost, the K-Computers system incurred significant development costs; thus, the procurement cost ratio was higher than for other supercomputers. Attention must be accorded to the interpretation of results as deflection may occur depending on the selection of the systems to be compared.

The limitations of this study are as follows. First, some major hardware factors such as I/O systems and storages were not reflected owing to lack of data. Therefore, additional factors should be further explored, investigated, and analyzed. We also did not consider the cost of the facilities in which a supercomputer is installed as part of the procurement cost. In addition, we did not consider other operating costs such as labor costs, software purchase costs, and other indirect expenses. Therefore, a more accurate model that reflects the general overall procurement and operation costs may be developed after collecting data from other supercomputer centers. On the other hand, weights were derived from the survey responses of experts in this study. We may consider the use of statistical methods such as factor analysis, principal component analysis, or other survey methods such as the Delphi method or the analytic hierarchy process. Finally, we compared the models presented in this paper only with the supercomputer performance measures provided by TOP500, but we plan to examine these models against various other measures such as Graph500 and I/O500 in future work.

Acknowledgement

This research was supported by the KISTI Program (No. K-19-L02-C05-S01), the EDISON Program through the National Research Foundation of Korea (NRF) (No. NRF-2011-0020576), the NRF and the Center for Women In Science, Engineering and Technology (WISSET-2019-233). A Grant was awarded by the Ministry of Science and ICT (MSIT) under the Program for Returners for R&D.

References

- [1] Gene H. Golub and James M. Ortega, *Scientific Computing and Differential Equations: An Introduction to Numerical Methods*, 1st Edition, Academic Press, Orlando, 1991.

- [2] E. Gallopoulos and A. Sameh, "CSE: content and product," *IEEE Computational Science and Engineering*, vol. 4, no. 2, pp. 39-43, June, 1997. [Article \(CrossRef Link\)](#)
- [3] PITAC (President's Information Technology Advisory Committee), *Computational Science: Ensuring America's Competitiveness*, the National Coordination Office for Information Technology Research and Development, Arlington, 2005.
- [4] National Research Council, *Getting Up to Speed: The Future of Supercomputing*, The National Academies Press, Washington, DC, 2005. [Article \(CrossRef Link\)](#)
- [5] Jack J. Dongarra, Piotr Luszczyk and Antoine Petit, "The LINPACK benchmark: past, present and future," *Concurrency and Computation: Practice and Experience*, vol. 15, no. 9, pp. 803-820, August, 2003. [Article \(CrossRef Link\)](#)
- [6] The TOP500 Supercomputer Sites. <https://www.top500.org>
- [7] National Academies of Sciences, Engineering, and Medicine, *Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020*, The National Academies Press, Washington, DC, 2016. [Article \(CrossRef Link\)](#)
- [8] Xiang Chen, Meizhuo Zhang, Minghui Wang, Wensheng Zhu, Kelly Cho and Heping Zhang, "Memory management in genome-wide association studies," *BMC Proceedings*, vol. 3, no. 7, pp. S54, December, 2009. [Article \(CrossRef Link\)](#)
- [9] Tzu-Yi Chen, Omid Khalili, Roy L. Campbell, Laura Carrington, Mustafa M. Tikir and Allan Snively, "Performance Prediction and Ranking of Supercomputers," *Advances in Computers*, Vol. 72, pp. 135-172, 2008. [Article \(CrossRef Link\)](#)
- [10] John L. Gustafson and Rajat Todi, "Conventional Benchmarks as a Sample of the Performance Spectrum," *Journal of Supercomputing*, pp 321-342, vol. 13, no. 3, May 1999. [Article \(CrossRef Link\)](#)
- [11] Ryusuke Egawa, Kazuhiko Komatsu, Yoko Isobe, Toshihiro Kato, Souya Fujimoto, Hiroyuki Takizawa, Akihiro Musa and Hiroaki Kobayashi, "Performance and Power Analysis of SX-ACE using HP-X Benchmark Programs," in *Proc. of IEEE Int. Conf. on Cluster Computing*, pp. 693-700, September 5-8, 2017. [Article \(CrossRef Link\)](#)
- [12] Vladimir Marjanović, José Gracia and Colin W. Glass, "HPC Benchmarking: Problem Size Matters," in *Proc. of Int. Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computing Systems*, pp. 1-10, November 14, 2016. [Article \(CrossRef Link\)](#)
- [13] Wu-chun Feng, Xizhou Feng and Rong Ge, "Green Supercomputing Comes of Age," *IT Professional*, vol. 10, no. 1, pp. 17-23, January, 2008. [Article \(CrossRef Link\)](#)
- [14] Jeffrey S. Vetter, *Contemporary High Performance Computing: From Petascale toward Exascale*, 1st Edition, Chapman and Hall, New York, 2013. [Article \(CrossRef Link\)](#)
- [15] Francesco Fraternali, Andrea Bartolini, Carlo Cavazzoni and Luca Benini, "Quantifying the Impact of Variability and Heterogeneity on the Energy Efficiency for a Next-Generation Ultra-Green Supercomputer," *IEEE Transactions on Parallel and Distributed Systems*, vol. 29, no. 7, pp. 1575-1588, July, 2018. [Article \(CrossRef Link\)](#)
- [16] Stephen W. Keckler, William J. Dally, Bruce Khailany, Michael Garland and David Glasco, "GPUs and the Future of Parallel Computing," *IEEE Micro*, vol. 31, no. 5, pp. 7-17, September, 2011. [Article \(CrossRef Link\)](#)
- [17] David Rohr, Sebastian Kalcher, Matthias Bach, Abdulqadir A. Alaqeeli, Hani M. Alzaidy, Dominic Eschweiler, Volker Lindenstruth, Sakhar B. Alkherefy, Ahmad Alharthy, Abdulelah Almubarak, Ibraheem Alqwaizy and Riman Bin Suliman, "An Energy-Efficient Multi-GPU Supercomputer," in *Proc. of IEEE Int. Conf. on High Performance Computing and Communications*, pp.42-45, August 20-22, 2014. [Article \(CrossRef Link\)](#)
- [18] David Rohr, Gvozden Neskovic and Volker Lindenstruth, "The L-CSC cluster: Optimizing power efficiency to become the greenest supercomputer in the world in the Green500 list of November 2014," *The Journal of Supercomputing Frontiers and Innovations*, vol. 2, no. 3, pp.41-48, July, 2015. [Article \(CrossRef Link\)](#)
- [19] Tapasya Patki, David K. Lowenthal, Barry Rountree, Martin Schulz, and Bronis R. de Supinski, "Exploring hardware overprovisioning in power-constrained, high performance computing," in *Proc. of Int. ACM Conf. on supercomputing*, pp.173-182, June 10-14, 2013.

- [Article \(CrossRef Link\)](#)
- [20] J.R. Neely, "The U.S. Government Role in HPC: Technology, Mission, and Policy," in *Proc. of Conf. on Comparing High Performance Computing in the U.S. and China*, pp.1-16, April 29-30, 2014.
 - [21] The Green500. <https://www.top500.org/green500>
 - [22] Everett Phillips and Massimiliano Fatica, "A CUDA Implementation of the High Performance Conjugate Gradient Benchmark," in *Proc. of High Performance Computing Systems. Performance Modeling, Benchmarking, and Simulation*, Springer International Publishing, pp. 68–84, 2015. [Article \(CrossRef Link\)](#)
 - [23] Jack Dongarra, Michael A Heroux and Piotr Luszczek, "High-performance conjugate-gradient benchmark: A new metric for ranking high-performance computing systems," *The International Journal of High Performance Computing Applications*, vol. 30, no. 1, pp. 3-10, February, 2016. [Article \(CrossRef Link\)](#)
 - [24] JaeHyuk Kwack and Gregory H. Bauer, "HPCG and HPGMG benchmark tests on multiple program, multiple data (MPMD) mode on Blue Waters—A Cray XE6/XK7 hybrid system," *Concurrency and Computation: Practice and Experience*, vol. 30, no. 1, October, 2017. [Article \(CrossRef Link\)](#)
 - [25] Joseph P. Martino, "A comparison of two composite measures of technology," *Technological Forecasting and Social Change*, vol. 44, no. 2, pp. 147-159, September, 1993. [Article \(CrossRef Link\)](#)
 - [26] Joseph P. Martino, *Technological Forecasting for Decision Making*, 3rd Edition, McGraw-Hill, New York, 1993.
 - [27] Shinobu Yoshimura, Muneo Hori and Makoto Ohsaki, *High-Performance Computing for Structural Mechanics and Earthquake/Tsunami Engineering*, 1st Edition, Springer International Publishing, Switzerland, 2016. [Article \(CrossRef Link\)](#)
 - [28] Thomas Sterling, Matthew Anderson and Maciej Brodowicz, *High Performance Computing – Modern Systems and Practices*, 1st Edition, Morgan Kaufmann, San Francisco, 2017. [Article \(CrossRef Link\)](#)
 - [29] Sun Yeong Heo, Duk Joon Chang, Jae Kyoung Shin, "Ordering Items from Ranking Procedures in Survey Research," *Survey Research*, vol. 9, no. 2, pp.29-49, July, 2008.
 - [30] Solmaz Salehian and Yonghong Yan, "Evaluation of Knight Landing High Bandwidth Memory for HPC Workloads," in *Proc. of Workshop on Irregular Applications: Architectures and Algorithms*, pp.10:1-10:4, November 12-17, 2017. [Article \(CrossRef Link\)](#)
 - [31] FUJITSU Supercomputer PRIMEHPC FX100. <http://www.fujitsu.com/global/products/computing/servers/supercomputer/primehpc-fx100/>
 - [32] Yuichiro Ajima, Takahiro Kawashima, Takayuki Okamoto, Naoyuki Shida, Kouichi Hirai, Toshiyuki Shimizu, Shinya Hiramoto, Yoshiro Ikeda, Takahide Yoshikawa, Kenji Uchida and Tomohiro Inoue, "The Tofu Interconnect D," in *Proc. of IEEE Int. Conf. on Cluster Computing*, pp.646-654, September 10-13, 2018. [Article \(CrossRef Link\)](#)
 - [33] Supercomputer at the Information Technology Center of Nagoya University. <http://www.icts.nagoya-u.ac.jp/en/sc/>
 - [34] Alfred P. Rovai, Jason D. Baker and Michael K. Ponton, *Social Science Research Design and Statistics: A Practitioner's Guide to Research Methods and IBM SPSS*, 2nd Edition, Watertree Press, Chesapeake, 2013.
 - [35] Kim Soon-Kwi, *Logistic regression analysis*, Kyowoo, Seoul, 2014.



Nam-Gyu Kim obtained his bachelor's degree and master's degree in computer science and engineering at Chung-Ang University in 2000 and 2002, respectively. And now he is a Ph.D. candidate in management of technology at Sungkyunkwan University. He has served as a senior researcher of Korea Institute of Science and Technology Information (KISTI) in Korea. He performed lots of SW R&D projects, including the national e-Science project, the AMGA development project, etc. In addition, he was the collaboration member of LIGO and KEK Belle projects as the site administrator to construct and manage the data grid site in Korea. Besides he participated in planning the second 5-year national supercomputing. At present, he participates in EDISON (ED-ucation-research Integration through Simulation On the Net) project, which is a web-based simulation service for education and research. He is interested in policies for HPC and computational science and engineering.



Noori On obtained her bachelor's degree and master's degree in management at Korea University of Technology and Education (KOREATECH) in 2007 and 2010, respectively. She worked as a Researcher at Korea Institute of Science and Technology Information (KISTI) during 5 years since 2010 and took part in making policies and plan of National Supercomputing. She has served as a researcher at KISTI since 2018. At present, she participates in EDISON (ED-ucation-research Integration through Simulation On the Net) project, which is a web-based simulation service for education and research. She is interested in policies for HPC and computational science and engineering.



Myoung-Ju Koh received the B.S. degree in Systems Management Engineering from Sungkyunkwan University (SKKU), in 2005 and M.S degree and Ph. D. degree in Industrial Engineering from SKKU, in 2007 and 2014, respectively. She worked in Science and Technology Policy Institute (STEPI) for six years from 2009. She has been serving as a Senior Researcher at Korea Institute of Science and Technology Information (KISTI) since 2015. She researches on research policy and strategic management of Korea National Supercomputing Center.



Jongsuk Ruth Lee received her Ph.D. in Computer Science from Univ. of Canterbury, New Zealand. She is a principal researcher at National Institute of Supercomputing and Networking, Korea Institute of Science and Technology Information (KISTI) and an adjunct faculty at Univ. of Science & Technology of Korea. Her research interests are smart learning, parallel/distributed computing & simulation, and big data handling



Keun-Tae Cho is Professor of Sungkyunkwan University with Management of Technology. He obtained his Ph.D. in Management of Technology from Sungkyunkwan University. He is currently studying, focusing on R&D Management, Management of Technology.