

Моделирование характеристик деятельности отраслевых и региональных подсистем

Экономико-математическое моделирование тенденций изменения ключевых показателей высокопроизводительных суперкомпьютеров*

Л.Е. ВАРШАВСКИЙ

Аннотация. В статье приводится экономико-математический анализ тенденций изменения показателей развития высокопроизводительных суперкомпьютеров. Рассматриваются разработанные автором модели динамики распространения высокопроизводительных суперкомпьютеров, а также структуры вычислительных мощностей. Исследуются экономико-математические модели «гонки» производительности суперкомпьютеров между отдельными странами. Даются оценки ключевых показателей суперкомпьютеров на кратко- и среднесрочный период.

Ключевые слова: суперкомпьютеры, микропроцессоры, динамика, производительность, энергоэффективность.

Введение

В настоящее время использование высокопроизводительных вычислений (high-performance computing, HPC) является неременным условием научно-технического развития. Важным сегментом рынка высокопроизводительных вычислений является быстро развивающийся рынок суперкомпьютеров, объем которого превышает 4 млрд. долл. [1].

Высокий спрос на суперкомпьютеры во многом обусловлен стремительным увеличением их производительности. Так, если в 1997 г. был превышен рубеж производительности суперкомпьютеров в 1 терафлопс, то в 2008 г. – 1,10 петафлопс (Пффлопс). Согласно рейтингу Top 500, с июня 2016 г. до настоящего времени (июнь 2017 г.) наиболее

производительным в мире является китайский суперкомпьютер Sunway TaihuLight, на котором достигнута скорость выполнения операций 93 Пффлопс.

Самые производительные суперкомпьютеры сконцентрированы в исследовательских центрах государственных структур и крупных компаний. Так, 59 из 117 суперкомпьютеров производительностью свыше 1 Петафлопс (Пффлопс)**, входивших в ноябре 2016 г. список 500 наиболее производительных суперкомпьютеров мира Top 500 (он публикуется 2 раза в год [2]), эксплуатировались в исследовательских центрах.

Однако суперкомпьютеры все более широко используются не только в исследовательском и

* Работа подготовлена при частичной финансовой поддержке Российского гуманитарного научного фонда (проект №15–02–00363).

** В списке Top 500 приводится ранжировка суперкомпьютеров по производительности в соответствии с тестом Linpack, связанным с решением линейных уравнений большой размерности.

академическом секторах, но и в бизнесе. Так, если в 1997 г. в Top 500, входил 161 суперкомпьютер из коммерческого сектора (industry), то в 2007 г. – уже 248. Примерно такое же число суперкомпьютеров из этого сектора было представлено и в ноябрьском списке Top 500 за 2016 г. (245 суперкомпьютеров). Драйверами активного применения суперкомпьютеров в бизнесе являются: рост затрат на создание и испытание образцов (прототипов) изделий; резкое снижение затрат на виртуальное моделирование и вычисления [2].

Следует отметить, что в последние 2 десятилетия «...рост производительности суперкомпьютеров происходит в равной мере, как за счет повышения быстродействия физических элементов, так и за счет увеличения числа элементов в вычислительной установке; последнее связано с развитием многоуровневого параллелизма обработки» [3, с. 578]. В связи с этим, актуальными являются исследование развития как физико-технологической базы суперкомпьютеров, основу которой составляют устройства микроэлектроники, так и их архитектуры.

В настоящей работе проводится экономико-математический анализ тенденций изменения ключевых показателей высокопроизводительных суперкомпьютеров, динамики их распространения и изменения структуры, а также факторов, влияющих на их энергоэффективность. Исследуются подходы к первоначальной оценке возможностей создания суперкомпьютера экзафлопсной производительности.

1. Анализ ключевых технико-экономических показателей оборудования для суперкомпьютеров

Динамика распространения высокопроизводительных суперкомпьютеров

Последние 2 десятилетия ознаменовались ускоренным технологическим развитием микроэлектроники и средств вычислительной техники. Существенно улучшились все основные параметры, характеризующие качественные и количественные характеристики микроэлектронных схем, долговременной памяти и создаваемых на их основе вычислительных устройств.

Развитие физико-технологической базы существенно повлияло на архитектуру и характеристики суперкомпьютеров. Так, в 1980–1990 гг. при построении суперкомпьютеров использовалась векторно-конвейерная структура, в рамках которой разные процессоры объединялись с общей памятью. Производительность систем составляла

тогда 1–10 Гигафлопс. В 1990–1995 гг. за счет применения массово-параллельных структур с распределенной памятью (MPP-MIMD), основанных на КМОП-технологии, удалось выйти на уровень производительности 100 Гигафлопс и более.

С середины 1990–х гг. построение суперкомпьютеров осуществлялось «...путем объединения крупносерийных микропроцессоров широкого применения» [4]. В 1997 г. был превышен рубеж производительности суперкомпьютеров в 1 терафлопс, в 2005 г. – 280 терафлопс, а в 2008 г. – 1,10 петафлопс (Пфлопс). В конце 2016 г. суммарная по показателю R_{max} производительность 500 суперкомпьютеров из списка Top 500 составила 672,1 Пфлопс.

Вместе с тем, необходимо отметить, что скорость освоения все более высокой производительности по показателю R_{max} суперкомпьютерами из списка Top 500 уменьшается. Так, уровень производительности в 1 терафлопс был достигнут всеми 500 суперкомпьютерами из этого списка в 2005 г., производительности в 100 терафлопс – в 2013 г., т.е. также за 8 лет. Однако за такой период времени, т.е. к концу 2016 г., уровень производительности в 1 Пфлопс был превышен лишь на 117 суперкомпьютерах. О динамике освоения высоких уровней производительности дает представление рис. 1, на котором приведена динамика доли f_t вычислительных мощностей суперкомпьютеров производительностью более 1 терафлопс, 100 терафлопс и 1 петафлопс в суммарной вычислительной мощности 500 суперкомпьютеров (в соответствии с показателем R_{max}).

Для количественной оценки освоения высоких уровней производительности суперкомпьютеров в настоящей работе использовалась логистическая зависимость,

$$df/dt=af(1-f), \tag{1}$$

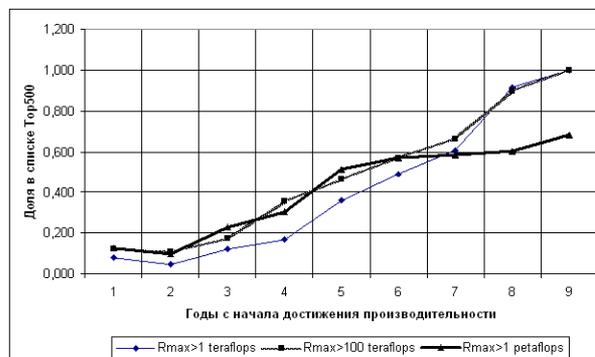


Рис. 1. Динамика освоения уровней производительности в 1 терафлопс, 100 терафлопс и 1 петафлопс суперкомпьютерами из списка Top 500

которая может быть приведена к следующему виду [5]:

$$\ln(f_t/(1-f_t)) = \alpha t + \beta, \quad (2)$$

где α характеризует скорость освоения уровней производительности суперкомпьютеров. Сопоставление оценок α этой зависимости для разных уровней производительности R_{max} свидетельствует о замедлении освоения достигаемых рекордных уровней (табл. 1; в скобках под значениями оценок здесь и далее приведены среднеквадратические ошибки оценок коэффициентов). Данная тенденция свидетельствует об отсутствии спроса или о неготовности широкого круга потребителей к эксплуатации высокопроизводительных компьютеров петафлопсного уровня.

Табл. 1

Оценки коэффициентов зависимости (1)

	$\alpha =$	$\beta =$	Коэффициент детерминации R^2
$R_{max} > 1$ teraflops	0,692 (0,088)	-3,975 (0,447)	0,911
$R_{max} > 100$ teraflops	0,581 (0,057)	-3,027 (0,287)	0,946
$R_{max} > 1$ petaflops	0,380 (0,047)	-2,378 (0,361)	0,905

Динамика производительности суперкомпьютеров

В настоящей работе проведен экономико-статистический анализ факторов, влияющих на производительность мощных суперкомпьютеров (применительно к суперкомпьютерам петафлопсного класса $R_{max} > 1$ Пфлопс). В табл. 2 представлены

коэффициенты корреляции производительности суперкомпьютеров и их основных технических характеристик ($R_{max} > 1$ Пфлопс) с 2010 по 2016 гг. Полученные данные подтверждают очевидное положительное влияние на производительность таких факторов, как электрическая мощность, количество ядер процессоров, а также ядер-ускорителей (сопроцессоров). Также обращает на себя внимание слабая корреляция показателей производительности и внутренней частоты процессоров. Такая ситуация во многом связана с различными подходами к архитектуре и составу оборудования, использованными в суперкомпьютерах, которые обусловлены спецификой решаемых задач.

По данным ноябрьского выпуска Top 500 за 2016 г. для выборки суперкомпьютеров производительностью свыше $R_{max} > 1$ Пфлопс построена следующая регрессионная зависимость, связывающая производительность (в Гигафлопс) с числом ядер PR_i и электрической мощностью W_e (в кВт):

$$R_{max} = 8,454 * PR + 184,243 * W_e + 1178011,376, \quad R^2 = 0,947. \quad (3)$$

(0,232) (70,904) (271092,052)

Расчеты, проведенные по данным за 2005–2016 гг., показывают более сильное влияние числа ядер на производительность суперкомпьютеров петафлопсного класса, чем на производительность менее мощных машин (так, эластичность этой группы компьютеров $\alpha = 1,110$ почти в полтора раза превосходит эластичность группы суперкомпьютеров производительностью $R_{max} > 100$ терафлопс, равную 0,739).

Динамика структуры суперкомпьютеров по их архитектуре

Наряду с анализом динамики производительности, значительный интерес представляет исследование динамики структуры суперкомпьютеров

Табл. 2

Динамика коэффициентов корреляции производительности суперкомпьютеров и их основных технических характеристик ($R_{max} > 1$ Пфлопс)

Годы	Производительность R_{max}	Электрическая мощность	Количество ядер процессоров	Количество ядер-ускорителей (сoproцессоров)	Внутренняя частота процессоров
2010	1,000	0,425	0,616	0,753	0,284
2011	1,000	0,899	0,984	-0,191	-0,482
2012	1,000	0,746	0,863	0,424	-0,348
2013	1,000	0,870	0,948	0,809	-0,255
2014	1,000	0,983	0,949	0,827	-0,188
2015	1,000	0,886	0,943	0,803	-0,239
2016	1,000	0,537	0,971	0,266	-0,312

по их архитектуре. В настоящее время в списке Top 500 присутствуют суперкомпьютеры с массивно-параллельной (Massively parallel processing – MPP) и кластерной (Cluster) архитектурой. Суперкомпьютеры с массивно-параллельной архитектурой MPP характеризуются в среднем более высокой производительностью, чем кластерные (до 2013 гг. они регулярно занимали 1–е место в рейтингах Top 500). Однако особенностью последнего десятилетия является уменьшение их доли в списке 500 наиболее производительных вычислительных машин. Так, доля суперкомпьютеров с архитектурой MPP в общем числе суперкомпьютеров уменьшилась с 20% в 2005 г. до 13,6% в 2016 г. Доля этих же суперкомпьютеров в суммарной производительности всех суперкомпьютеров из списка Top 500 снизилась за тот же период с почти 50 до 41%.

В настоящей статье исследовалась динамика структуры суперкомпьютеров по их архитектуре. С этой целью по данным за 2008–2016 гг. была построена следующая динамическая модель конкуренции в форме векторной авторегрессии (VAR):

$$\begin{pmatrix} \Delta \text{LN}(R_{\max t}^{\text{Cluster}}) \\ \Delta \text{LN}(R_{\max t}^{\text{MPP}}) \end{pmatrix} = \begin{pmatrix} \phi_1 & \theta_1 \\ \phi_2 & \theta_2 \end{pmatrix} \begin{pmatrix} \text{LN}(R_{\max t-1}^{\text{Cluster}}) \\ \text{LN}(R_{\max t-1}^{\text{MPP}}) \end{pmatrix} + \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}, \quad (4)$$

где $\text{LN}(R_{\max t}^{\text{Cluster}})$, $\text{LN}(R_{\max t}^{\text{MPP}})$ — натуральные логарифмы суммарной производительности совокупности суперкомпьютеров с архитектурой Cluster и MPP; Δ — знак приращения во времени; ϕ_i, θ_i, d_i — постоянные коэффициенты. Параметры построенной модели представлены в табл. 3.

Система (4) имеет устойчивое положение равновесия, характеризуемое координатами: $\text{LN}(R_{\max t}^{\text{Cluster}}) = 8,717$ и $\text{LN}(R_{\max t}^{\text{MPP}}) = 8,149$, что соответствует производительности суперкомпьютеров с архитектурой Cluster и MPP $R_{\max t}^{\text{Cluster}} = 6104,4$ Пфлопс и $R_{\max t}^{\text{MPP}} = 3460,3$ Пфлопс. Положению равновесия соответствуют доли архитектур Cluster и MPP в совокупной производительности, равные соответственно 63,8 и 36,2%. В 2016 г. эти доли составляли 58,9 и 41,1%.

Расчеты на основе модели (4) показывают, что при сохранении существующих тенденций, в 2020 г. доля мощных суперкомпьютеров архитектуры Cluster составит 62,6%, а архитектуры MPP – 37,4% (рис. 2). При этом суммарная производительность совокупности из 500 наиболее производительных суперкомпьютеров превысит 1,65 экзафлопс.

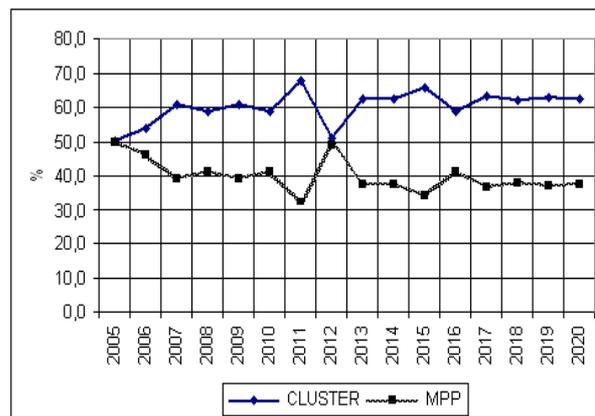


Рис. 2. Доля суперкомпьютеров с архитектурой MPP и Cluster в суммарной производительности всех суперкомпьютеров

Динамика показателей энергоэффективности суперкомпьютеров

Происходящее повышение производительности суперкомпьютеров, как правило, сопровождается ростом их электрической мощности. Так, суммарная электрическая мощность, потребляемая всеми суперкомпьютерами из списка Top 500 устойчиво возрастала с 161 МВт в 2007 г. до 696 МВт в 2016 г. (по данным ноябрьских выпусков Green 500) [6].

Несмотря на существенный прогресс в повышении энергоэффективности, энергоёмкость высокопроизводительных суперкомпьютеров остается высокой. В частности, согласно рейтингу Top 500, проведенному в ноябре 2016 г., наиболее производительный в мире китайский суперкомпьютер

Табл. 3

Оценки параметров ϕ_i, θ_i, d_i модели (4), $i=1,2$

Совокупность суперкомпьютеров с архитектурой	ϕ_i	θ_i	d_i	Коэффициент детерминации R^2
Cluster	-0,103 (0,037)	0,000 (0,000)	0,894 (0,149)	0,523
MPP	-1,362 (0,720)	1,225 (0,522)	0,421 (3,652)	0,656

Sunway TaihuLight 93,0 Пфлопс, потребляет около 15,37 МВт эл. мощности. Второй по производительности суперкомпьютер Tianhe-2 (MilkyWay-2) с быстродействием 33,9 Пфлопс имеет электрическую мощность 17,81 МВт, а необходимая электрическая мощность для третьего по производительности суперкомпьютера Titan, имеющего быстродействие 17,59 Пфлопс, также составляет немалую величину – 8,2 МВт [6,7].

Следует также отметить, что из-за высокой плотности размещения вычислительных мощностей в ограниченном пространстве, проблема отвода тепла в суперкомпьютерах стоит более остро, чем в центрах обработки данных (ЦОД). В настоящее время, по мнению специалистов, наилучшие характеристики, как и для ЦОД, имеет технология водяного охлаждения, которая позволяет снизить показатель PUE*** с 1,5–2 до 1,05-1,06 [8, 9].

В связи с высокой энергоемкостью суперкомпьютеров, начиная с 2007 г. регулярно публикуются списки Green 500, в которых проводится ранжировка наиболее производительных суперкомпьютеров по величине соотношения между их производительностью R_{max} и используемой электрической мощностью. В соответствии с ноябрьским за 2016 г. списком Green 500, максимальная удельная производительность суперкомпьютеров составляла 9462.1 Мегафлопс/Вт. В тоже время, как показывает анализ приведенных в этом списке данных, у 46% суперкомпьютеров из 500 самых мощных, показатель энергоэффективности составляет менее 10% от максимального уровня. Следует отметить, что по уровню энергоэффективности отечественные суперкомпьютеры долгие годы почти на порядок уступали лучшим зарубежным образцам, хотя в последнее время ситуация улучшается (в 2016 г. у нас достигнут уровень 2056,6 Мегафлопс/Вт, что соответствует 112 месту в списке Green 500) [6].

В связи с высокой энергоемкостью современных суперкомпьютеров, представляет интерес экономико-статистический анализ факторов, влияющих на их удельную производительность (т.е. производительность в расчете на 1 Вт электрической мощности). В настоящей работе такой анализ проведен в статике (применительно к суперкомпьютерам петафлопсной производительности $R_{max} > 1$ Пфлопс) и в динамике (применительно к наиболее производительным в 2005-2014 гг. суперкомпьютерам).

Так, в табл. 4 представлены коэффициенты корреляции удельной производительности суперкомпьютеров и их основных технических характеристик ($R_{max} > 1$ Пфлопс) с 2010 по 2016 гг. Полученные данные показывают, что влияние на удельную производительность исследовавшихся факторов является существенно менее значимым, чем на производительность, и неоднаправленным (табл. 4, ср. с табл. 2).

Вместе с тем, из физических соображений ясно, что в целом определяющее влияние на производительность и электрическую мощность суперкомпьютеров должны оказывать количество ядер и тактовая частота процессоров, а также топологический размер схем. Так, на основе выборки, составленной из наиболее производительных в 2016 г. суперкомпьютеров (с производительностью $R_{max} > 1$ Пфлопс), была получена следующая регрессионная зависимость между удельной производительностью

$$\frac{R_{max}}{W_e} \text{ в Мфлопс/Вт и числом ядер на 1 Вт } \frac{PR}{W_e} :$$

$$\ln\left(\frac{R_{max}}{W_e}\right) = 0,761 * \ln\left(\frac{PR}{W_e}\right) + 4,078, \quad R^2 = 0,608 \quad (5)$$

(0,239) (0,057)

(включение большего числа факторов в эту и приводимые ниже регрессионные зависимости приво-

Табл. 4

Динамика коэффициентов корреляции удельной производительности суперкомпьютеров (Мегафлопс/Вт) и их основных технических характеристик ($R_{max} > 1$ Пфлопс)

	Удельная производительность Мфлопс/Вт	Число ядер на 1 Вт	Доля ядер-ускорителей (со-процессоров) в общем количестве ядер	Внутренняя частота процессоров
2010	1,000	0,660	0,279	0,589
2012	1,000	0,906	-0,202	-0,744
2014	1,000	0,648	0,309	-0,402
2016	1,000	0,415	0,439	-0,192

*** PUE (англ. power usage effectiveness) – отношение полной подведенной мощности к реально затраченной на работу вычислительных систем)

дит к незначимым с позиций t-статистик дополнительным коэффициентам).

Для представительной выборки показателей суперкомпьютеров петафлопсного класса ($R_{max} > 1$ Пфлопс), содержащихся в ноябрьских выпусках Top 500 за 2013–2015 гг., получены следующие регрессионные зависимости между удельной энергоем-

костью 1 ядра процессора $\frac{W_e}{PR}$ в Вт/ядро и тактовой частотой FR в МГц:

$$2013 \text{ г.: } \ln\left(\frac{W_e}{PR}\right) = 2,404 * \ln(FR) - 22,852, \quad R^2 = 0,613; \quad (6a)$$

(0,354) (2,744)

$$2014 \text{ г.: } \ln\left(\frac{W_e}{PR}\right) = 2,751 * \ln(FR) - 25,430, \quad R^2 = 0,448; \quad (6b)$$

(0,441) (3,425)

$$2015 \text{ г.: } \ln\left(\frac{W_e}{PR}\right) = 2,380 * \ln(FR) - 15,798, \quad R^2 = 0,637. \quad (6c)$$

(0,268) (2,079)

Как следует из (6a), (6c), значения коэффициентов полученных зависимостей при логарифмах составляют 2,380 и 2,751, что согласуется с данными о том, что потребляемая процессорами электрическая мощность зависит от тактовой частоты в степени 2–2,5 [10]. Именно в связи с этим, в последнее десятилетие одним из важных направлений повышения энергоэффективности суперкомпьютеров стало увеличение не тактовой частоты, а числа процессоров и ядер.

В тоже время приведенный в работе [10] статистический анализ динамических рядов показывает, что для выборки из наиболее производительных суперкомпьютеров с процессорами на основе схем разного топологического размера, показатель $R_{max t}$ в млрд. Мфлопс связан следующей функциональной зависимостью с числом ядер процессоров PR_t и топологическим размером чипов hp_t в нм:

$$\ln R_{max t} = 10,940 + 0,852 * \ln(PR_t / hp_t^2), \quad (7a)$$

а необходимая для работы суперкомпьютера электрическая мощность $W_{e t}$ (в кВт) оказывается связанной с отношением PR_t / hp_t зависимостью:

$$\ln W_{e t} = 4,476 + 0,443 * \ln(PR_t / hp_t). \quad (7b)$$

Из последних зависимостей следует, в частности, что рост удельной производительности

$\frac{R_{max}}{W_e}$ наиболее производительных суперкомпьютеров был во многом связан с увеличением числа ядер процессоров и уменьшением топологического размера микросхем.

2. Моделирование «гонки» производительности суперкомпьютеров между отдельными странами

Особенностью текущего десятилетия стала утрата единоличного лидерства США в производительности суперкомпьютеров. Так, в 2010 г. наиболее производительным суперкомпьютером из списка Top 500 стал китайский суперкомпьютер NUDT производительностью 2,566 Пфлопс. В 2011 г. на первое место вышел японский суперкомпьютер K computer производительностью 10,51 Пфлопс. После короткого периода лидерства США в 2012 г. (суперкомпьютер Titan производительностью 17,59 Пфлопс) на первое место опять вышли китайские суперкомпьютеры Tianhe-2 (MilkyWay-2, с 2013 г. по 2015 гг. включительно), а с 2016 г. – суперкомпьютер Sunway, производительность которого более, чем в 5 раз выше, чем у Titan. В результате, в конце 2016 г. общая производительность суперкомпьютеров США из списка Top 500 по показателю R_{max} была на 50 Пфлопс меньше, чем суммарная производительность суперкомпьютеров Китая и Японии (соответственно 228 Пфлопс против 278 Пфлопс).

Следует отметить, что создание производительных суперкомпьютеров в Китае осуществляется в рамках пятилетнего плана, с целью преодоления отставания в уровне экономического развития от уровня развитых стран, а также с целью обеспечения собственных разработок в области высоких технологий. Последнее вызывает озабоченность в США, так как в стремлении к обладанию Китаем собственными высокими технологиями там усматривается угроза повышения военного потенциала этой наиболее населенной страны мира [12]. Вместе с тем, нельзя исключать того, что в ближайшем будущем лидерство в области наиболее производительного суперкомпьютера опять на некоторое время перейдет к американцам, которые реализуют ряд проектов, связанных с созданием эксафлопсного суперкомпьютера.

Таким образом, текущее десятилетие отмечено началом «гонки» производительности суперкомпьютеров между ведущими странами мира. В связи с этим, представляет практический интерес межстрановой анализ перспектив развития суперкомпьютерного потенциала.

С этой целью в настоящей работе построена модель «гонки» производительности суперкомпьютеров между США с одной стороны и Китаем и Японией с другой в форме векторной авторегрессии (VAR):

$$\begin{pmatrix} \ln(R_{max t}^{USA}) \\ \ln(R_{max t}^{China+Japan}) \end{pmatrix} = \begin{pmatrix} \lambda_1 & \mu_1 \\ \lambda_2 & \mu_2 \end{pmatrix} \begin{pmatrix} \ln(R_{max t-1}^{USA}) \\ \ln(R_{max t-1}^{China+Japan}) \end{pmatrix} + \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}, \quad (8)$$

где $R_{max t}^{USA}$, $R_{max t}^{China+Japan}$ – суммарная производительность совокупности суперкомпьютеров США, а также Китая и Японии; $\lambda_i, \mu_i, g_i, i=1,2$ – постоянные коэффициенты. Параметры построенной модели представлены в табл. 5.

Расчеты на основе модели (8) показывают, что при сохранении существующих тенденций, в 2020 г. суммарная производительность мощных суперкомпьютеров Китая и Японии будет превышать соответствующий показатель в США на 32% (в настоящее время величина этого превышения составляет почти 22%) (рис. 3).

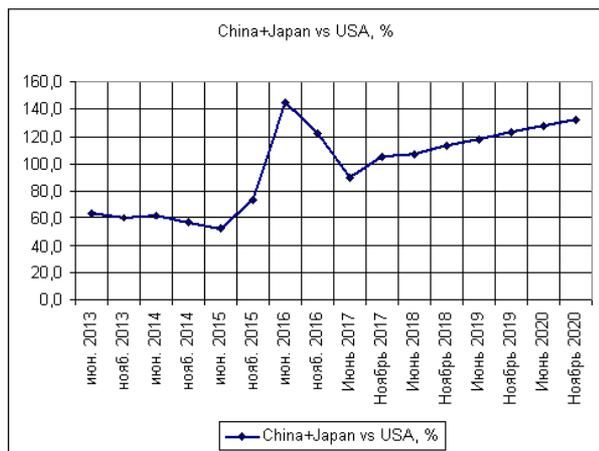


Рис. 3. Динамика соотношения между суммарной производительностью всех суперкомпьютеров Китая и Японии с одной стороны, и США с другой, %

3. Проблема создания суперкомпьютера экзафлопсной производительности

Бурный рост производительности суперкомпьютеров стимулировал возрастание спроса на высокопроизводительные вычисления НРС со стороны исследовательского сектора, государственных структур и бизнеса. В связи с этим, перед специалистами ряда высокоразвитых стран, в частности, в США, Японии и Китае,

поставлена проблема создания суперкомпьютеров экзафлопсной производительности (т.е. производительностью более 10^{18} флоп в секунду). При этом необходимо иметь в виду, что в настоящее время не существует принципиальных технических ограничений для создания суперкомпьютеров такого класса на основе существующего оборудования [13]. Однако, несмотря на достигнутые в последние годы результаты в области энергоэффективности вычислительных устройств, при наилучшем в настоящее время (первая половина 2017 г.) уровне энергоэффективности для устройств петафлопсного класса (9,462 Гигафлопс/Вт для суперкомпьютера DGX SaturnVc производительностью $R_{max}=3,307$ Пфлопс), необходимая электрическая мощность экзафлопсного суперкомпьютера составила бы 105,7 МВт. Этой мощности достаточно для обеспечения электроэнергией крупного городского квартала с населением несколько десятков тысяч человек. При полной загрузке его годовые расходы на электроэнергию составили бы порядка 70–80 млн. долл (в условиях США).

Поэтому, например, в США, перед разработчиками ставится задача создания экзафлопсного суперкомпьютера электрической мощности порядка 20 МВт. Так, с целью ограничения роста масштабов энергопотребления, а также повышения его эффективности, агентством DARPA Министерства обороны США была в 2010 г. разработана программа «Вездесущие высокоэффективные вычисления»(Ubiquitous High Performance Computing – UHPC). В рамках этой программы ставилась задача обеспечить размещение в одном кабинете компьютерные мощности в 1 Пфлопс с энергоэффективностью 50 Гигафлопс/Вт (это более, чем в 5 раз больше текущего уровня). Предполагалось, что первый образец системы УНРС будет создан в 2018 г. [14].

Для первоначальной оценки реальности решения такой проблемы можно использовать зависимости (7a), (7b). Расчеты с использованием этих зависимостей показывают, что при сохранении

Табл. 5

Оценки параметров λ_i, μ_i, g_i модели (8), $i=1,2$

Совокупность суперкомпьютеров с архитектурой	λ_i	μ_i	g_i	Коэффициент детерминации R^2
США	0,664 (0,121)	0,194 (0,078)	0,913 (0,264)	0,982
Китай+ Япония	1,368 (0,131)	0,000 (0,000)	-1,917 (0,550)	0,893

ческих процессоров GPU (главным образом, GPU NVIDIA), а также сопроцессоров Intel Xeon Phi и ряда других производителей.

В целом полученные оценки, по-видимому, свидетельствуют о реальности создания суперкомпьютера производительностью 1 экзафлопс и электрической мощностью 20–40 МВт при переходе на производство схем с малыми топологическими размерами. Возможные сроки такого перехода, сопряженного с высокими затратами, зависят, в частности, от прогресса в создании: экономически эффективных установок экстремальной ультрафиолетовой фотолитографии (EUV); быстродействующих памяти и межсоединений; нового поколения массивно-параллельных систем; программного обеспечения, в частности, программных средств обнаружения и устранения ошибок, которые могут возникнуть ввиду повышения риска более частых неполадок из-за резкого увеличения (до миллиардов) числа компонентов в вычислительных системах (ensuring correctness in the face of faults; resilience at scale) и др. [15]. Одно из перспективных направлений ускорения перехода к энергоэффективным суперкомпьютерам состоит в создании мультиархитектурных вычислительных систем, в которых «...объединяются наборы процессоров для прикладных программ с различными формами параллелизма и набором функционально специализированных процессоров» [10].

Важно подчеркнуть, что в конечном счете перспективы развития суперкомпьютеров связаны с успешным решением проблемы как финансирования, так и снижения стоимости. Роль показателя стоимости особенно высока ввиду как дороговизны высокопроизводительных суперкомпьютеров, так и относительно малой продолжительности их жизненного цикла по сравнению с другими высокотехнологичными объектами сопоставимой стоимости. Так, из 7 эксплуатировавшихся в 2010 г. высокопроизводительных суперкомпьютеров ($R_{max} > 1$ Пфлопс), к 2016 г. в списках Top 500 и Green 500 осталось всего 3. Наглядным примером может служить судьба суперкомпьютера Roadrunner в Лос-Аламосской национальной лаборатории, который первым превзошел петафлопсный рубеж производительности в 2008 г. Эксплуатация этого суперкомпьютера стоимостью 121 млн. долл. была прекращена в конце марта 2013 г., т.е. менее, чем через 5 лет после ввода в строй [16].

Стоимость введенных в эксплуатацию несколькими годами позже, а также разрабатываемых высокопроизводительных машин на порядок выше, чем у Roadrunner. Так, по некоторым данным стоимость первого в списке Top 500 в ноябре

2011 г. японского суперкомпьютера K Computer превысила миллиард долларов [13]. По последним данным, стоимость одной экзафлопсной системы, разрабатываемой одновременно в США и Китае, составит 300–500 млн. долл. Кроме того, до ввода первого образца потребуется ежегодно не менее 1 млрд. долл. для проведения необходимых исследований и разработок. Желаемый объем финансирования суперкомпьютера нового поколения экзафлопсной производительности, который намечается создать в Японии, также составлял порядка 1 млрд. долл. (11 млрд. иен) [17–19].

Вместе с тем, в новом варианте Проекта создания экзафлопсного суперкомпьютера в США (U.S. Exascale Computing Project) предусмотрено создание первого такого американского суперкомпьютера в 2021 г. В Китае первый экземпляр экзафлопсного суперкомпьютера предполагается создать в 2020 г. [19]. Следует отметить, что намеченные сроки создания экзафлопсных систем близки к оценкам, полученным в настоящей статье.

Заключение

Разработанные в статье модели позволяют провести анализ динамики распространения высокопроизводительных суперкомпьютеров, их производительности, изменения структуры, а также факторов, влияющих на энергоэффективность. Они могут быть также использованы при первоначальной оценке перспектив достижимости устанавливаемых на будущее уровней производительности и эффективности суперкомпьютеров.

Полученные оценки, по-видимому, свидетельствуют о реальности создания суперкомпьютера производительностью 1 экзафлопс и электрической мощностью 20–40 МВт при переходе на производство схем с меньшими топологическими размерами в ближайшие 5–7 лет.

Литература

1. <https://www.hpcwire.com/2017/04/06/supercomputer-sales-drove-2016-hpc-market-record-11-2-billion/> доступ 15.04.2017.
2. <http://www.top500.org/> доступ 02.06.2017.
3. Turek D. The Age of the Supercomputer. <http://www.businessweek.com/stories/2007-05-07/the-age-of-the-supercomputerbusinessweek-business-news-stock-market-and-financial-advice/> доступ 05.06.2013.
4. Левин В.К. Высокопроизводительные вычислительные системы—суперкомпьютеры в России// Базовые лекции по электронике. Том II. Твердо-

- тельная электроника/Под общ. ред. В.М.Пролейко. М.: Техносфера. 2009. С.576–594.
5. *Сахал Д.* Технический прогресс: концепции, модели, оценки. М: Финансы и статистика. 1985.
 6. <http://www.green500.org/lists.php>/доступ .
 7. <http://www.cnews.ru/news/top/index.shtml?2012/06/27/494386>/доступ 02.07.2012.
 8. *Левшин И.* Из жизни алгоритмов. HPC Superstar // Суперкомпьютеры. № 13. 2013. С. 10–12.
 9. <http://efytimes.com/e1/99599/Eurotech-Supercomputer-At-CINECA-Breaks-World-Record-For-Energy-Efficiency>/доступ 20.06.2013.
 10. *Митропольский Ю.И.* Элементная база и архитектура будущих суперкомпьютеров// Микроэлектроника. 2015. Том44. №3. С. 163–179.
 11. *Варшавский Л.Е.* Проблемы повышения энергоэффективности аппаратных средств в области информационных технологий//Труды ИСА РАН. Т. 63. Вып. 3. 2013. С. 3-19.
 12. *Feldman M.* The DOE and NSA Construct Doomsday Scenario for American HPC. <https://www.top500.org/news/the-doe-and-nsa-construct-doomsday-scenario-for-american-hpc/> доступ 07.04.2017.
 13. *Левшин И.* Из жизни звезд. // Суперкомпьютеры. № 13. 2013. С. 6–9.
 14. *Feldman M.* DARPA Sets Ubiquitous HPC Program in Motion. http://www.hpcwire.com/hpcwire/2010-08-10/darpa_sets_ubiquitous_hpc_program_in_motion.html/доступ 19.04.2011.
 15. *Reed D. and Dongarra J.* Exascale Computing and Big Data//Communications of the ACM, 2015, v. 58, No 7, pp. 56–68.
 16. <http://www.bbc.co.uk/news/technology-21993132>/доступ 03.06.2015.
 17. <http://efytimes.com/e1/fullnewsr.asp?edid=105870>/доступ 21.05.2014.
 18. <http://fcw.com/articles/2013/07/17/exoflop-supercomputing.aspx>/доступ 21.07.2013.
 19. <https://www.hpcwire.com/2017/04/26/messina-update-u-s-path-exascale-15-slides/> доступ 21.05.2017.

Варшавский Леонид Евгеньевич. Заведующий лабораторией ИСА ФИЦ ИУ РАН. Главный научный сотрудник ЦЭМИ РАН. Окончил МИФИ в 1970 году. Доктор экономических наук. Количество печатных работ: 120. Область научных интересов: математическое моделирование рыночных процессов и прогнозирование показателей рынков высокотехнологичной и капиталоемкой продукции, математическое моделирование инвестиционных и инновационных стратегий промышленных фирм, математическое моделирование макроэкономических процессов, методы управления динамическими системами, экономика науки. E-mail: hodvar@mail.ru

Analysis and modeling trends of key indicators of high-performance supercomputers

L.E.Varshavsky

Abstract. Economico-statistical analysis of indicators of high-performance supercomputers is carried out. Models of diffusion of high-performance supercomputers, of dynamics of their structure and of influence of key factors on their performance are proposed. Approaches to initial assessment of possibility and terms of creating exaflops supercomputers with the help of built models based on analysis of current trends in energy efficiency and possible progress in microelectronics are proposed.

Keywords: *supercomputers, microprocessors, dynamics, productivity, energy efficiency.*

References

1. <https://www.hpcwire.com/2017/04/06/supercomputer-sales-drove-2016-hpc-market-record-11-2-billion/> (accessed April 15, 2017).
2. <http://www.top500.org/> (accessed June 2, 2017).
3. *Turek D.* The Age of the Supercomputer. <http://www.businessweek.com/stories/2007-05-07/the-age-of-the-supercomputerbusinessweek-business-news-stock-market-and-financial-advice/> (accessed June 5, 2013).
4. *Levin V.K.* Vysokoproizvoditel'nye vychislitel'nye sistemy—superkomp'yutery v Rossii [High performance computing systems—supercomputers in Russia] // Tom II. Tverdotel'naya elektronika [Foundations and History of Development // Basic Lectures on Electronics. Vol. II. Solidstate Electronics] /Pod obshch. red. V.M.Proleiko. Moscow: Tekhnosfera, 2009, pp. 576–594.
5. *Sahal D.* Patterns of Technological Innovation. Addison-Wesley Publishing Co., Reading, 1980.
6. <http://www.green500.org/lists.php> (accessed June 2, 2017).
7. <http://www.cnews.ru/news/top/index.shtml?2012/06/27/494386/> (accessed July 2, 2012).
8. *Levshin I.* Iz zhizni algoritmov. NRS Superstar [From life of algorithms]. Superkomp'yutery=Supercomputers, № 13, 2013, c. 10–12.
9. <http://efytimes.com/e1/99599/Eurotech-Supercomputer-At-CINECA-Breaks-World-Record-For-Energy-Efficiency/> (accessed June 20, 2013)..
10. *Mitropol'skij Ju.I.* Jelementnaja baza i arhitektura budushhih superkomp'yuteroi [The element base and architecture of future supercomputers]. Mikr ojelektronika=Microelectronics. 2015, vol 44, no 3, pp. 163–179.
11. *Varshavskii L.E.* Problemy povysheniya energoeffektivnosti apparatnykh sredstv v oblasti informatsionnykh tekhnologii [Problems of Energy Efficiency Increase of Information Technologies Equipment Infrastructure]. Trudy ISA RAN=Proceedings of ISA RAS, 2013, vol. 63, no. 3, pp. 3-19.
12. *Feldman M.* The DOE and NSA Construct Doomsday Scenario for American HPC. <https://www.top500.org/news/the-doe-and-nsa-construct-doomsday-scenario-for-american-hpc/> (accessed April 7, 2017).
13. *Levshin I.* Iz zhizni zvezd [From life of stars]. Superkomp'yutery=Supercomputers, № 13, 2013, c. 6–9.
14. *Feldman M.* DARPA Sets Ubiquitous HPC Program in Motion. http://www.hpcwire.com/hpcwire/2010-08-10/darpa_sets_ubiquitous_hpc_program_in_motion.html (accessed Apr 19, 2011).
15. *Reed D. and Dongarra J.* Exascale Computing and Big Data//Communications of the ACM, 2015, v. 58, No 7, pp. 56–68.
16. <http://www.bbc.co.uk/news/technology-21993132> (accessed June 3, 2015).
17. <http://efytimes.com/e1/fullnewsp.asp?edid=105870/> (accessed May 21, 2014).
18. <http://fcw.com/articles/2013/07/17/exoflop-supercomputing.aspx> (accessed July 21, 2013).
19. <https://www.hpcwire.com/2017/04/26/messina-update-u-s-path-exascale-15-slides/> (accessed May 21 2017).

Varshavsky Leonid. Head of the laboratory of the ISA FRC CSC RAS, chief scientist of the CEMI RAS. Graduated from MEPHI in 1970. Author of 120 articles and monographs. Fields of scientific interests: mathematical modeling of market processes and forecasting market indicators of high technology and capital intensive products, mathematical modeling of investment and innovative strategies of firms, mathematical modeling of macroeconomic processes, methods of control theory, economics of science.

E– mail: hodvar@mail.ru