International Journal of
Science and Engineering Investigations
vol. 8, issue 86, March 2019
Received on February 28, 2019
ISSN: 2251-8843

Prediction Studies of Electricity Use of Global Computing in 2030

Anders S. G. Andrae
Huawei Technologies Sweden AB, Skalholtsgatan 9, 16494 Kista, Sweden
(anders.andrae@huawei.com)

Abstract-The electricity use of the information technology (IT) sector - consisting of demand from computing, transmission and production - is of large interest. Here a theoretical framework describing how the total global electricity demand - associated with the computing instructions done in servers and computers - is used to estimate the electricity use in 2030. The proposed theoretical framework is based on the following parameters: instructions per second, joules per transistor, and transistors per instruction as well as a distinction between general and special purpose computing. The potential predictions - made possible with the proposed equations - include the electricity used by the data centres based on utilization of the processors therein and estimations of the electricity use of the processors used in fixed and mobile networks and end-user devices. Production of IT equipment/hardware is excluded as well as transmission in mobile and core networks use stage. Predictions for computing 2030 vary a lot from 1 to 4487 TWh depending on which transistor technology will be dominant handling the instructions. Two other prediction techniques - based on instructions per joule and joules per operation - give similar results.

Keywords- *Computing, Computations, Data centres, E-factor, Electricity, Instructions, Operations, Power use, Processing, Servers, Switching, Transistors*

I. Introduction

The World is demanding more and more energy and especially electrical energy will be used in many more applications tomorrow than today. For instance, electric vehicles [1,2] and sustainable hydrogen production [3] for fuel cell vehicles - or reformed internal combustion engines - means extra demand of electric power [4].

The digitalization of most businesses will also require additional amounts [5]. Some [9] argue that a "tsunami" of data is about to be unleashed according to Fig. 1.

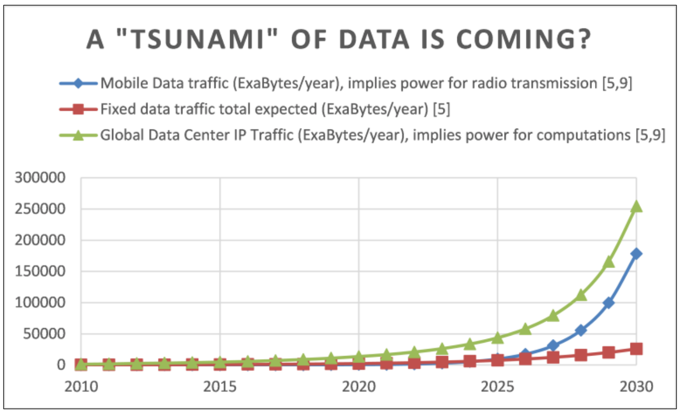

Figure 1. Possible evolution of total global data traffic toward 2030 [5,9].

An earlier prediction model for the entire information technology (IT) sector [5] includes the fact that energy efficient smartphones and tablets might be used instead of desktops and laptops for streaming videos. It is suspected therefore that total consumer devices use stage power declines until 2030 despite being affected by the same transistor technology problems as the rest of the computing sector. There is also a lot of logical reasoning/speculation around "Smart IT" which can optimize the energy efficiency (e.g. in buildings) and save perhaps more per year than IT uses itself. So far the total global electricity use is increasing [6], however slower than ITs own electricity use [5,7]. Electricity saving is happening, but also more demand and consumption for e.g. the IT infrastructure. Some argue that the cloud services are being responsible for a rebound effect [8] and that can play a role in this dilemma. It is important to find better ways of estimating and forecasting the global electricity use by computing as the next decade will present itself with many challenges. Although trend analyses [5,9] might be reasonable enough, it is worthwhile checking other approaches estimating computing power. The IT Sector is divided into computing, transmission and production. The present study refines the computing share as far as electric power predictions.

II. PROBLEM FORMULATION

In the present prediction a new methodological framework [10] is used for estimating a major share of the global power use of IT. In the present estimate the hypothesis is that computing instructions done in servers will - in around 2030 - under certain circumstances use several thousands of TWh, unless breakthroughs in semiconductor technologies are reached in the next decade. The result (order of magnitude) is thought to be comparable to previous estimations such as [7]. To the authors knowledge the global electricity use of computing instructions – based on instructions per joule and instructions per second, and General Purpose Computing (GPC) and Special Purpose Computing (SPC) – has yet to be estimated for 2030.

III. PROBLEM SOLUTION

In this prediction an approach – (1)-(5) - is used for extrapolating the potential electricity use of current and future computing instructions, $\left(\frac{J}{s}\right)_Y$.

$$\left(\frac{J}{s}\right)_Y = \frac{\left(\frac{Trans}{Ins}\right)_Y \times \left(\left(\frac{Ins}{s}\right)_{GPC,Y} \times \left(\frac{J}{Trans}\right)_{GPC,Y} + \left(\frac{Ins}{s}\right)_{SPC,Y} \times \left(\frac{J}{Trans}\right)_{SPC,Y}\right)}{CPU_{utilization}} \quad (1)$$

where

$\left(\frac{J}{s}\right)_Y$ = potential electricity use of computing instructions in data centres and computers in year Y.

$\left(\frac{Trans}{Ins}\right)_Y$ = transistors per instruction in year Y.

$\left(\frac{Ins}{s}\right)_{GPC,Y}$ = instructions per second for GPC in year Y (average global traffic).

$\left(\frac{J}{Trans}\right)_{GPC,Y}$ = Joules per transistor for GPC in year Y.

$\left(\frac{Ins}{s}\right)_{SPC,Y}$ = instructions per second for SPC in year Y (average global traffic).

$\left(\frac{J}{Trans}\right)_{SPC,Y}$ = Joules per transistor for SPC in year Y.

$CPU_{utilization,Y}$= the average share of a data centre's - or a computers - total power use that is used by the Central Processing Units (CPUs) – and Graphical Processing Units (GPUs) - in year Y.

Values for each parameter needed to quantify $\left(\frac{J}{s}\right)_Y$ will be estimated for chosen years between Y = 2007 and Y = 2030.

A. *Electricity per instruction – switching electricity*

$\left(\frac{J}{Trans}\right)_{SPC,Y}$, the switching electricity [11] in year Y of special purpose computing, SPC is given by (2).

$$\left(\frac{J}{Trans}\right)_{SPC,Y} = Ef_Y \times k_B \times T = Ef_Y \times 1.38 \times 10^{-23} \times T \quad (2)$$

$\left(\frac{J}{Trans}\right)_{SPC,DS,Y}$, the switching electricity in year Y of SPC for "dark silicon" transistors is given by (3).

$$\left(\frac{J}{Trans}\right)_{SPC,DS,Y} = Ef_{DS,Y} \times k_B \times T = Ef_{DS,Y} \times 1.38 \times 10^{-23} \times T \quad (3)$$

where

k_B= Boltzmann's constant [J/K].

T= Temperature at which the transistor is operating [K].

Ef_Y= energy/enthropy factor (e-factor) in year Y.

$Ef_{DS,Y}$= energy/enthropy factor (e-factor) for "dark silicon" transistors in year Y.

T is assumed to be 313.73 Kelvin to be consistent with 11] and International Technology Roadmap for Semiconductors Roadmap ITRS) [11] for $\left(\frac{J}{Trans}\right)_{SPC,2030} = 1aJ$

1) *Parameter estimations*

Table I lists the evolution of e-factors and switching energies from 1989 to 2018 and predictions for 2030 based on roadmaps from [11], [12], [13], [14] and [15].

TABLE I. EVOLUTION AND EXTRAPOLATION OF E-FACTORS AND SWITCHING ELECTRICITIES FROM 1989 TO 2030 FOR FIVE DIFFERENT ROADMAPS

Year Y	Ef_Y	$\left(\frac{J}{Trans}\right)_{SPC,Y}$ (zeptoJoule = $10^{-21}\left(\frac{J}{Trans}\right)$)	$\left(\frac{J}{Trans}\right)_{SPC,DS,Y}$ dark silicon, zeptoJoule
1989	9932000[16]	41137803 [11]	
2003	78500 [16]	325143 [11]	
2006	67700 [16]	560819 [11]	
2007	57483 [11]	250000 [11]	
2008	18900 [16]	82200 (2)	156566 [11]
2012	4500 [16]	19600 (2)	74555 [11]
2014	1750 [16]	7610 (2)	28994 [11]
2015	1500 [16]	6520 (2)	24852 [11]
2017			13889 [11]
2018			10256 [11]
2030 ITRS roadmap [11]	230 (2)	1000 [11]	3330
2030 Frank roadmap [12]	100 [16]	435	1448
2030 Bennett roadmap [13]	40 [16]	174	579
2030 Landuaer Roadmap [14]	0.7 (2)	3 [11]	9.99
2030, Reversible Computing roadmap [15]	0.04 [16]	0.17 (2)	0.67

A slow-down of the improvement of $\frac{J}{Trans}$ is obvious from 2008. Between 1989 and 2007 $\left(\frac{J}{Trans}\right)_{SPC,Y}$ improved around 25% per year, and from 2008 to 2015 by 30% per year. $\left(\frac{J}{Trans}\right)_{SPC,DS,Y}$ only improved around 23% per year from 2008 to 2015. With such a deceleration it is questionable if even the ITRS roadmap, being the least optimistic, can be realized.

Based on the average ratio (3.33) between $\left(\frac{J}{Trans}\right)_{SPC,DS,Y}$ and $\left(\frac{J}{Trans}\right)_{SPC,Y}$ – the "dark silicon effect" - for 2008, 2012, 2014 and 2015, $\left(\frac{J}{Trans}\right)_{SPC,DS,Y}$ for 2030 can be estimated for different roadmaps.

B. Instructions per second

The only source found mentioning a roadmap for global traffic expressed as $\left(\frac{Ins}{s}\right)_{Y}$ is [17].

$\left(\frac{Ins}{s}\right)_{Y}$ will affect $\left(\frac{J}{s}\right)_{Y}$ tremendously. Table II shows the evolution of $\left(\frac{Ins}{s}\right)_{Y}$ for selected years.

TABLE II. EVOLUTION OF INSTRUCTIONS PER SECOND

Year Y	GPC, (Zetta$\left(\frac{Ins}{s}\right)$=$10^{21}\left(\frac{Ins}{s}\right)$	SPC, (Zetta$\left(\frac{Ins}{s}\right)$=$10^{21}\left(\frac{Ins}{s}\right)$
1986	2.86×10^{-7} [17]	4.19×10^{-7} [17]
2007	6.24×10^{-3} [17]	0.185 [17]
2012	0.035	1.47
2030	18 [17]	2570 [17]

$\left(\frac{Ins}{s}\right)_{GPC,2012}$ and $\left(\frac{Ins}{s}\right)_{SPC,2012}$ in Table II are obtained from extrapolation between 2007 and 2030.

C. Transistors per instruction

$\left(\frac{Trans}{Ins}\right)_{Y}$ in (4) refers to the transistors in a chip which are involved in doing an instruction. $\left(\frac{Trans}{Ins}\right)_{Y}$ is the most difficult parameter to estimate in the present prediction model.

$$\left(\frac{Trans}{Ins}\right)_{Y} = \frac{\left(\frac{J}{s}\right)_{servers,Y}}{\left(\left(\frac{Ins}{s}\right)_{GPC,Y}\times\left(\frac{J}{Trans}\right)_{GPC,Y}+\left(\frac{Ins}{s}\right)_{SPC,Y}\times\left(\frac{J}{Trans}\right)_{SPC,Y}\right)} \quad (4)$$

The assumed difference between GPC and SPC as far as switching electricity is given by (5).

$$\left(\frac{J}{Trans}\right)_{GPC,Y} = \left(\frac{J}{Trans}\right)_{SPC,Y}\times 100 \quad (5)$$

It has been estimated that servers globally in 2007 used 82 TWh [18], i.e. $\left(\frac{J}{s}\right)_{servers,2007} = 9.31\, GW$ and 105 TWh in 2012 [18], i.e. $\left(\frac{J}{s}\right)_{servers,2012} = 11.98\, GW$. Hence $\left(\frac{Trans}{Ins}\right)_{2007}$ and $\left(\frac{Trans}{Ins}\right)_{2012}$ can be estimated as shown in (6) and (7) from values in Tables I and II:

$$\left(\frac{Trans}{Ins}\right)_{2007} = \frac{\left(\frac{J}{s}\right)_{servers,2007}}{\left(\left(\frac{Ins}{s}\right)_{GPC,2007}\times\left(\frac{J}{Trans}\right)_{GPC,2007}+\left(\frac{Ins}{s}\right)_{SPC,2007}\times\left(\frac{J}{Trans}\right)_{SPC,2007}\right)} = \frac{9.31\times10^{9}}{7.37\times10^{18}\times2.5\times10^{-14}+2.18\times10^{20}\times2.50\times10^{-16}} = 38994 \quad (6)$$

$$\left(\frac{Trans}{Ins}\right)_{2012} = \frac{\left(\frac{J}{s}\right)_{servers,2012}}{\left(\left(\frac{Ins}{s}\right)_{GPC,2012}\times\left(\frac{J}{Trans}\right)_{GPC,DY,2012}+\left(\frac{Ins}{s}\right)_{SPC,2012}\times\left(\frac{J}{Trans}\right)_{SPC,DY,2012}\right)} = \frac{11.98\times10^{9}}{0.0352\times10^{21}\times74555\times10^{-19}+1.47\times10^{21}\times74555\times10^{-21}} = 32331 \quad (7)$$

As shown by (6)-(7), $\left(\frac{Trans}{Ins}\right)_{Y}$ is not constant over the years. For the baseline it is assumed that $\left(\frac{Trans}{Ins}\right)_{2012}$ is constant until 2030. As shown in Table I, the decrease of $\frac{J}{Trans}$ started to slow down from 2008, and including this fact between 2008 and 2012 will affect the value of $\left(\frac{Trans}{Ins}\right)_{Y}$. Table III shows the assumed evolution of $\left(\frac{Trans}{Ins}\right)_{Y}$.

TABLE III. ASSUMED EVOLUTION OF TRANSISTORS PER INSTRUCTION 2007 TO 2030.

Year Y	$\left(\frac{Trans}{Ins}\right)_{Y}$
2007	38994
2012	32331
2030	*32331*

D. Other methods by which computing power can be estimated

Next follows two additional approaches by which computing power use can be determined and predicted.

1) Operations Per Second per Watt and Joules per Operation

Yet another way to estimate the global electricity use of computation operations is according to (8) which can employ available Operation Per Second per milliwatt and Joules per Operation roadmaps [19].

$$\left(\frac{J}{year}\right)_{Y} = \frac{1}{\left(\frac{\left(\frac{Ops,chip\ P\ nm\ node}{s}\right)_{Y}}{\left(\frac{J}{s}\right)_{Y,chip\ P\ nm\ node}}\right)}\times\left(\frac{s}{year}\right)\times\left(\frac{Ops}{s}\right)_{Y} \quad (8)$$

where

$\left(\frac{J}{year}\right)_{Y}$ = Average electricity use in year Y of computing instructions.

$\left(\frac{Ops,chip\ P\ nm\ node}{s}\right)_{Y}$ = Operations performed - in the typical average chip $P\ nm\ node$ - per second in year Y.

$\left(\frac{J}{s}\right)_{Y,chip\ P\ nm\ node}$ = Power consumption of typical average chip $P\ nm\ node$ in year Y.

$\left(\frac{s}{year}\right)$= Average seconds per year.

$\left(\frac{Ops}{s}\right)_Y$ = Operations per second (global average traffic) in year Y, assumed identical to $\left(\frac{Ins}{s}\right)_{GPC,Y}$ + $\left(\frac{Ins}{s}\right)_{SPC,Y}$.

$\left(\frac{Trans}{Ins}\right)_Y$ is absent from (8) which is an advantage. In section V.A some results will be presented using (8).

2) Combine instructions per joule with instructions per second

Data centre estimations – such as [5,7,9] - for 2030 can be "tested" with an approach based on instructions per second (IPS) [17] and instructions per joule (IPJ) [20].

As shown in Table IV, if Koomey's Law (9) [20] for electricity efficiency improvement - 55% per year since 2000 - holds to 2030, and we have 3.0×10^{24} IPS by then - we still will have several thousands of TWh needed for computing. This is true even if the CPU share of the electricity use of the data centres and computers would be 100% (i.e. no extra electric power needed for cooling, fans, uninterruptable power supply etc.) and if no split – regarding energy efficiency - is made between SPC and GPC. Nowadays the CPU share of the electricity use of the data centres might be just 20% for some smaller data centres [21] however much higher - like 70% - for professional hyperscale data centres [22].

Equations (9) to (12) show how the electricity use of computing can be estimated.

$$\left(\frac{Ins}{kWh}\right)_Y = e^{(0.4401939\times Y-849.1617)} \tag{9}$$

$$\left(\frac{Ins}{J}\right)_Y = \frac{\left(\frac{Ins}{kWh}\right)_Y}{3.6\times10^6} \tag{10}$$

$$\left(\frac{J}{s}\right)_Y = \frac{\left(\frac{Ins}{s}\right)_{,Y}}{\left(\frac{Ins}{J}\right)_Y} \tag{11}$$

$$\left(\frac{J}{s}\right)_Y = \frac{\left(\frac{Ins}{s}\right)_{GPC,Y}}{\left(\frac{Ins}{J}\right)_{GPC,Y}} + \frac{\left(\frac{Ins}{s}\right)_{SPC,Y}}{\left(\frac{Ins}{J}\right)_{SPC,Y}} \tag{12}$$

where

$\left(\frac{Ins}{J}\right)_{GPC,Y}$ = IPJ for General Purpose Computing in year Y

$\left(\frac{Ins}{J}\right)_{SPC,Y}$ = IPJ for Special Purpose Computing in year Y

$\left(\frac{Ins}{kWh}\right)_Y$ = Instructions per kWh for Computing in year Y.

The electricity use will be estimated for the year 2030 for SPC and GPC and compared to the values for 2030 in [5] and the present method based on switching electricity roadmaps. Table IV results only refers to computing related electricity in the data centre and computers, which is far from reality. This means that the consumption would be higher if other sources of power use than computing would be included.

TABLE IV. 2030 ELECTRICITY USE OF COMPUTING BASED ON INSTRUCTIONS PER KILOWATTHOUR AND INSTRUCTIONS PER SECOND

	$\left(\frac{Ins}{s}\right)_{GPC,Y}$	$\left(\frac{Ins}{s}\right)_{SPC,Y}$	$\left(\frac{Ins}{kWh}\right)_Y$	$\left(\frac{Ins}{J}\right)_{GPC,Y}$	$\left(\frac{Ins}{J}\right)_{GPC,Y}$	$\left(\frac{J}{s}\right)_Y$	TWh
Y = 2030	2.12×10^{22} [17]	3.03×10^{24} [17]	1.98×10^{19} (9)	5.5×10^{10} (5)	5.5×10^{12} (10)	9.42×10^{11} (12)	8253

IV. RESULTS

Here follows the calculation for the ITRS Roadmap [11] – with $\left(\frac{J}{Trans}\right)_{SPC,DS,2030,ITRS}$ according to (1):

$\left(\frac{J}{s}\right)_{ITRS\ Roadmap,2030} = \frac{32331\times(2.13\times10^{22}\times3.33\times10^{-16}+3.06\times10^{24}\times3.33\times10^{-18})}{1} = 558.77\times10^9 W = 4887\ TWh$.

Here follows the calculation for the Frank Roadmap [12] – with $\left(\frac{J}{Trans}\right)_{SPC,DS,2030,Frank}$ according to (1):

$\left(\frac{J}{s}\right)_{Frank\ Roadmap,2030} = \frac{32331\times(2.13\times10^{22}\times1.45\times10^{-16}+3.06\times10^{24}\times1.45\times10^{-18})}{1} = 243\times10^9 W = 2125\ TWh.$

Here follows the calculation for the Bennett Roadmap [13] – with $\left(\frac{J}{Trans}\right)_{\boldsymbol{SPC,DS,2030,Bennett}}$ according to (1):

$\left(\frac{J}{s}\right)_{Bennett\ Roadmap,2030} = \frac{32331\times(2.13\times10^{22}\times5.79\times10^{-17}+3.06\times10^{24}\times5.79\times10^{-19})}{1} = 97\times10^9 W = 850\ TWh.$

Here follows the calculation for the Landauer Roadmap [14] – with $\left(\frac{J}{Trans}\right)_{\boldsymbol{SPC,DS,2030,Landauer}}$ according to (1):

$\left(\frac{J}{s}\right)_{ICT,Landauer\ Roadmap,2030} = \frac{32331\times(2.13\times10^{22}\times1.01\times10^{-18}+3.06\times10^{24}\times1.01\times10^{-20})}{1} = 1.7\times10^9 W = 15\ TWh.$

Here follows the calculation for the Reversible Computing Roadmap [15] – with $\left(\frac{J}{Trans}\right)_{SPC,DS,2030,Reversible\ Computing}$ according to (1):

$$\left(\frac{J}{s}\right)_{Reversible\ Computing\ Roadmap,2030} = \frac{32331\times(2.13\times10^{22}\times5.79\times10^{-20}+3.06\times10^{24}\times5.79\times10^{-22})}{1} = 0.097\times10^{9}W = 0.8\ TWh.$$

The uncertainty range is huge for all parameters. For the Landauer and Reversible Computing Roadmaps the electricity use will be insignificant.

Fig. 2 shows the main results for the roadmaps [12-15] compared to previous estimations for $\left(\frac{J}{s}\right)_{2015}$ and $\left(\frac{J}{s}\right)_{2030}$ expressed as TWh. The results for all roadmaps assumes $CPU_{utilization,2030} = 1$ which might be a little overoptimistic.

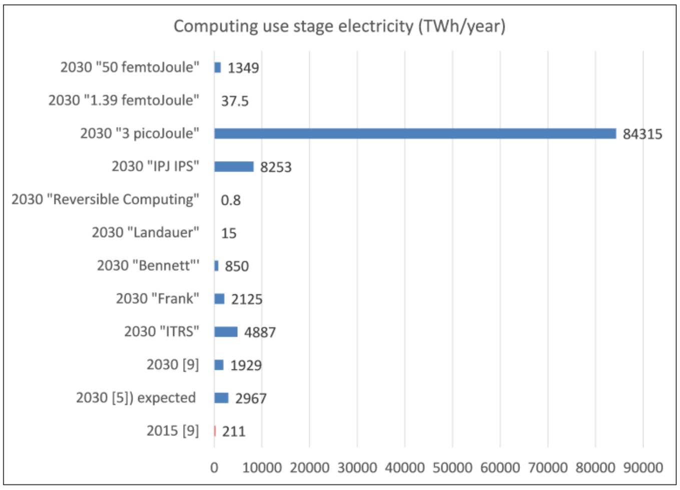

Figure 2. Comparison of roadmaps prediction results for 2030 with other predictions for data centres

Fig. 2 shows some similarities between earlier predictions [5,7,9] for data center computing power and the present roadmap prediction.

V. DISCUSSION

The present projections about global electricity usage of computing are still in line with the current situation. However, the absolute TWhrs are probably lower for 2015 TWhrs but same order of magnitude. We are in an optimistic trajectory along the "best case scenario" as outlined in [5] until around 2023 where the TWhrs will start to rise more or less rapidly. We are depending ourselves so much on the digitalization that – in one scenario – other electricity consuming sectors will have to stand back so data centres and networks can have the available power.

Three trends which strengthen the hypothesis of more power consumption from IT are: the exponential data (no matter how data is defined) demand of existing services such as video streaming [9], long change cycles in fundamental technologies, and new unforeseen demands of data. If the surging data demand is a foregone conclusion – which it probably is – then there could be other users of power which will have less access to it. Pricing of power and data will anyway probably solve the power issues if they become severe.

Moreover, it is not evident whether Artificial Intelligence (AI) will drive or reduce electricity use globally overall. However, AI deep neural networks will probably generate more data – and instructions - and thereby drive more electricity consumption.

Specifically for the present predictions there are several sources of important uncertainties. Two of the most important are for $\left(\frac{Trans}{Ins}\right)_Y$ and global data traffic $\left(\frac{Ins}{s}\right)_Y$ as there are no real roadmaps for those parameters. $\left(\frac{Ins}{s}\right)_Y$ is based on the number of cores in chips used globally so it has potential to be forecasted. If there will be 2.5 Yotta instructions per second (average global traffic) in 2030 – as predicted [17] and the ITRS roadmap [11] will be followed - there will be several thousands TWh extra electricity consumption, even with $CPU_{utilization,2030} = 1$. Less than 100% utilization will of course mean more electricity use.

Anyway, 100% utilization is perhaps plausible – especially in some parts of the World - as the most efficient run data centres have already set a tone for those less efficient and that has saved - and will save - power overall. There is an awareness in the industry which has helped slow-down the electricity use along the "best case scenario" as described in [5] so far.

A. *A breakthrough is necessary*

There needs to be a massive breakthrough for chips if the electricity should not rise in ICT. The Bennett roadmap [13] represents such a breakthrough, let alone the Landuaer [14] and Reversible Computing Roadmaps [15,16].

The presently excluded production of IT hardware is in 2030 expected to require several hundred TWhrs [5], adding to those factors which underestimate the present results. Mobile data transmission is another driver for overall IT power demand [5,9].

With current trends - and what seems possible in the next decade – technical research is necessary but not enough to prevent a high rise in the next decade.

There also need to be sharp implementations if the rapid rise is to be prevented.

The Bennett roadmap is similar to a prediction [19] listing two orders of magnitude improvements for energy per operation between 2000 and 2020. Energy per operation is likely similar to the inverse of instructions per joule [16]. Using (8) from section II.D.1, two examples are derived using data from [16] and [19]:

$$\left(\frac{J}{year}\right)_{2030,3.1\ pJ} = \frac{1}{\left(\frac{5.76\times10^{13}}{180}\right)} \times 31.536 \times 10^{6} \times 3.08 \times 10^{24} = 3.03 \times \frac{10^{20}\ J}{year} \rightarrow 84315\frac{TWh}{year}$$

This example means that if a “traditional” 180 W processor chip using 5 nm node [16] – 3.1 pJ/operation - would process the anticipated operations in 2030 with current transistor technology, an absurd amount of electricity will be used for computing.

The same chip using reversible computing [16] would use only 0.08 W and 1.39 fJ/operation:

$$\left(\frac{J}{year}\right)_{2030,1.39\ fJ} = \frac{1}{\left(\frac{5.76\times10^{13}}{0.08}\right)} \times 31.536 \times 10^{6} \times 3.08 \times 10^{24} = 1.35 \times \frac{10^{17}\ J}{year} \rightarrow 37.5\frac{TWh}{year}$$

Moreover, using other data from Table 1.2 in [19] for nanoelectronics with 1 fJ/operation:

$$\left(\frac{J}{year}\right)_{2030,\ 1\ fJ} = \frac{1}{\left(\frac{600\times10^{6}}{0.6\times10^{-6}}\right)} \times 31.536 \times 10^{6} \times 3.08 \times 10^{24} = 9.71 \times 10^{16}\ J \rightarrow 26.95\ TWh$$

If the breakthroughs on chip level – similar to [12] or [13] - have not been reached before the data “tsunami” have hit, there will be some interesting dilemmas in the power sector and society.

Pangrle [23] discussed a projected efficiency in 2023 of 1300 million operations per second per Watt being 19 times short of a 40 MW target for a specific data centre/super computer. Obviously, energy efficiencies of 769 (1300 million operations per second per Watt) and 40 pJ/operation (24.7 billion operations per second per Watt) will not be sufficient seven years later in 2030 if by then the global traffic is 3 Yotta operations/instructions per second. Something like 50 fJ/operation, i.e. 20 [Terraoperations/s]/[Watt] will be required to keep the power consumption in data centres under a certain control in 2030.

VI. CONCLUSION

A framework based on transistor physics is used to predict the 2030 electricity use of computing. The range of potential outcomes of electricity usages for computing at large is enormous, actually several orders of magnitude depending on which roadmap will prevail. Based on the predictions done with the present method, the electricity use of the computing infrastructure will in 2030 reach several thousands extra TWh. The reason is simply that the rate of energy efficiency improvements will not likely keep up with the rate of computing instructions.

VII. NEXT STEPS

Obviously many more permutations are necessary. Many new values of all parameters in the present framework need to be derived and collected. The electricity demand surge might be delayed into the 2030s thanks to quantum energy efficient nanochips [24]. Moreover, it need to be investigated what is the power consumption of current commercial quantum computers [25]. In a nutshell, the e-factor for quantum computing should be estimated. The verification of possible contradictions between total number of operations and total number of instructions would be worthwhile. The electric power use (e.g. TWhrs) obtained - from operations per year multiplied with energy per operation - should be compared more carefully to the present estimates.

REFERENCES

[1] H. Moon, S.Y. Park, C. Jeong, and J. Lee. “Forecasting electricity demand of electric vehicles by analyzing consumers ’ charging patterns,” Transportation Research Part D: Transport and Environment, Vol. 62, pp. 64-79, 2018. DOI: https://doi.org/10.1016/j.trd.2018.02.009

[2] P. Ryan. “Electricity Demand and Implications of Electric Vehicle and Battery Storage Adoption”. In Transition Towards 100% Renewable Energy, Springer, Cham. 2018, pp. 391-398. DOI: https://doi.org/10.1007/978-3-319-69844-1_35

[3] I. Vincent and D. Bessarabov. “Low cost hydrogen production by anion exchange membrane electrolysis: a review,” Renewable and Sustainable Energy Reviews, Vol. 81, No. 2, pp. 1690-1704, 2018. DOI: https://doi.org/10.1016/j.rser.2017.05.258

[4] A.S.G. Andrae. “Drawing the fresco of the electricity use of information technology in 2030 – Part II,” Researchgate.net., 2019. DOI: 10.13140/RG.2.2.31813.91361

[5] A.S.G. Andrae and T. Edler. ”On global electricity usage of communication technology: trends to 2030,” Challenges, Vol. 6, No. 1, pp. 117-157, 2015. DOI: https://doi.org/10.3390/challe6010117

[6] V. Petit. “The Transition to a New Electric World,” In The New World of Utilities, Springer, Cham. 2019. pp. 5-32. DOI: https://doi.org/10.1007/978-3-030-00187-2_2

[7] T. Masuhara. “The Future of Low-Power Electronics,” In CHIPS 2020, Vol. 2, B. Hoefflinger, Ed. Springer, Cham., 2016, pp. 21-50. DOI: https://doi.org/10.1007/978-3-319-22093-2_2

[8] H.J. Walnum and A.S.G. Andrae. “The internet: explaining ICT service demand in light of cloud computing technologies,” In Rethinking Climate and Energy Policies, T. Santarius, H.J. Walnum and C. Aall, Eds. Springer, Cham., 2016, pp. 227-241. DOI: https://doi.org/10.1007/978-3-319-38807-6_13

[9] A.S.G. Andrae. “Projecting the chiaroscuro of the electricity use of communication and computing from 2018 to 2030,” Researchgate.net. 2019. DOI: 10.13140/RG.2.2.25103.02724

[10] A.S.G. Andrae. “Drawing the fresco of the electricity use of information technology – Part I,” Researchgate.net. 2019. DOI: 10.13140/RG.2.2.18392.14080

[11] J.L. Summers. “From ZettaBytes to zeptoJoules ‐ can digital demand outstrip the physical limits?,” Presentation held at RISE SICS Open House Kista, Sweden 19 April 2018.

[12] M.P. Frank. “Approaching the physical limits of computing,” In Proceedings of 35th International Symposium on Multiple-Valued Logic (ISMVL´05). IEEE, 19-21 May 2005, Calgary, BC, Canada, pp. 168-185. DOI: 10.1109/ISMVL.2005.9

[13] C.H. Bennett. “The thermodynamics of computation ‒ a review, “ International Journal of Theoretical Physics, Vol. 21, No. 12, pp. 905-940, 1982. DOI: https://doi.org/10.1007/BF02084158

[14] R. Landauer. “Dissipation and noise immunity in computation and communication,” Nature, Vol. 335, No. 6193, pp. 779-784, 1988. DOI: https://doi.org/10.1038/335779a0

[15] W. Porod. “The Thermodynamics of Computation: A Contradiction,” In Energy Limits in Computation, C.S. Lent, A.O. Orlov, W. Porod and G.L. Snider, Eds. Springer, Cham., 2019. pp. 141-154. DOI: https://doi.org/10.1007/978-3-319-93458-7_4

[16] J.L. Summers. "From ZettaBytes to zeptoJoules – will digital demand outstrip the physical limits?," at Data Centre World, 15-16 March, 2017, London, United Kingdom.

[17] Z.W. Xu. "Cloud-sea computing systems: Towards thousand-fold improvement in performance per watt for the coming zettabyte era," Journal of Computer Science and Technology, Vol. 29, No. 2, pp. 177-181, 2014. DOI: https://doi.org/10.1007/s11390-014-1420-2

[18] W. Van Heddeghem, S. Lambert, B. Lannoo, D. Colle, M. Pickavet, and P. Demeester. "Trends in worldwide ICT electricity consumption from 2007 to 2012, " Computer Communications, Vol. 50, pp. 64-76, 2014. DOI: https://doi.org/10.1016/j.comcom.2014.02.008

[19] B. Hoefflinger. News on Eight Chip Technologies. In CHIPS 2020, Vol. 2, B. Hoefflinger, Ed. Springer, Cham., 2016, pp. 1-19. DOI: https://doi.org/10.1007/978-3-319-22093-2_1

[20] J. Koomey, S. Berard, M. Sanchez, and H. Wong. "Implications of historical trends in the electrical efficiency of computing," IEEE Annals of the History of Computing, Vol. 33, No. 3, pp. 46-54, 2011. DOI: 10.1109/MAHC.2010.28

[21] R. Bashroush. "A Comprehensive Reasoning Framework for Hardware Refresh in Data Centres," IEEE Transactions on Sustainable Computing, Vol. 3, No. 4, pp. 209-220, 2018. DOI: 10.1109/TSUSC.2018.2795465

[22] A. Shehabi, S.J. Smith, E. Masanet, J. Koomey. "Data center growth in the United States: decoupling the demand for services from electricity use," Environmental Research Letters, Vol. 13, No. 12, 2018, p. 124030. DOI: https://doi.org/10.1088/1748-9326/aaec9c

[23] B. Pangrle. "News on energy-efficient large-scale computing," In CHIPS 2020, Vol. 2, B. Hoefflinger, Ed. Springer, Cham., 2016, pp. 165-170. DOI: https://doi.org/10.1007/978-3-319-22093-2_9

[24] D.A Gangloff, G. Ethier-Majcher, C. Lang, E.V. Denning, J.H. Bodey, D.M. Jackson et al. "Quantum interface of an electron and a nuclear ensemble,", Science 21 Feb. 2019. DOI: 10.1126/science.aaw2906.

[25] D. Castelvecchi. IBM's quantum cloud computer goes commercial. Nature News, Vol. 543, No. 7644, 2017, 159. DOI: 10.1038/nature.2017.21585

How to Cite this Article:

Andrae, A. S. (2019). Prediction Studies of Electricity Use of Global Computing in 2030. International Journal of Science and Engineering Investigations (IJSEI), 8(86), 27-33. http://www.ijsei.com/papers/ijsei-88619-04.pdf