Data Processing Technology Forecast

Dr. Lawrence G. Roberts
Advanced Research Projects Agency

During the past decade, computer technology has improved rapidly, reducing the cost of computation power and increasing the maximum capability available. Users' requirements for increased capability are simultaneously increasing, necessitating the regular replacement of computer equipment at each installation. The purpose of this paper is to develop as formal a model as possible of the cost trade-off in a computer replacement cycle and extract guidance from this analysis on the optimal computer replacement strategy.

The average rate of growth of computer technology was determined by analyzing 39 large scientific computers. The rate of growth was used to determine the hardware costs (for both leasing and purchasing) as a function of:

1. the number of years before acquisition of a new computer,
2. the number of years the old computer is retained after delivery of the new machine and
3. the factor by which the installation's computer workload increases each year

The software cost incurred by the conversion of programs from the old to the new computer was added to obtain the composite hardware-software cost function. That cost function was minimized, and the following results were obtained:

1. For both lease and purchase, full overlap (the retention of old computer so that there are always two computers available) is the most desirable policy.

2. If the equipment is leased and computers are fully overlapped, the optimal number of years before the delivery of the next computer is always less than 1.5 years.

3. If the equipment is purchased, the follow-on system should be obtained in 2.5 – 3.5 years. However, the cost does not rise rapidly after the minimum is past.

4. The purchase is more desirable than lease when the workload growth rate is low or when the software conversion is high.

Performance Measure

In order to be specific about the rate of change taking place, it is necessary to define measures of performance and cost performance. Procurement practice has come to utilize benchmark tests to measure system throughput. Even though there are many inaccuracies associated with designing a benchmark series, it is still the best method available. Therefore, in order to assign a numerical performance measure to each variety of computer, a measure was chosen which matched as closely as possible the performance ratios obtained in a number of benchmark tests; megabits per second (MBS) flowing into the CPU(s). This measure is superior to simply measuring the average instruction rate since long word length machines are capable of doing more per instruction. Also, for some new machine organizations, the instruction rate is irrelevant; e.g., CDC's STAR design where a streaming operation initiated with one instruction is capable of producing thousands of results. Megabits per second is calculated for standard computers by multiplying the instruction rate by the data word length.

If the maximum instruction rate is used, the MBS value obtained correlates very well with throughput ratios observed from benchmark tests. However, some machines do not operate at full efficiency and, in those cases, the average instruction rate and, thus, the MBS have been reduced to correspond to the benchmark data. Since benchmarks differ, any single measure is, at best, an average. It is felt, however, that the MBS compiled herein are certainly within a factor of two and probably much closer when compare to observed throughput rates.

On the attached graph of throughput speed, it can be seen that the speed of computers introduced in any given year varies considerably. This is because slower, less expensive computers are introduced after the technology develops. The average growth rate of the ensemble of scientific computers was determined by computing the least squares best fit exponential. The growth rate thus determined is 1.85 per year:

$$S = 1.85\ (t - 59)$$

Cost – Performance Measure

Cost-performance can now be measured as the ratio of the throughput rate (MBS) to the computer system cost. This results in the measure “bits per second per dollar”. System cost is a much less sharply defined quantity than the throughput rate. For example, the central processor on a machine may cost $1M, the memory $.5 – 1.5M and the I/O equipment $.5 – 2.5M. Thus, the system cost may vary between $2M and $5M, making it hard to define the average case. The costs used herein are based on the complete system cost for a well balanced utilization of the CPU, but not including the cost associated with extensive peripheral equipment such as satellite computers and display consoles. The attached graph of throughput per dollar shows that the machines are tightly grouped together for this measure. The best fit exponential in this case has a slower growth rate, 1.565 per year:

$$s/c = 1.565\ (t - 59.4)$$

Almost all computers are introduced within two years of this average curve. The reason for this may be due to manufacturers’ marketing policy, since a computer far below this curve would not be financially worth introducing.

Relationship of Speed to Cost

In order to determine the hardware cost for a particular procurement where the computer speed required is the variable, it is necessary to find the function of speed and cost which has the minimal variation at each point and time. This relationship was found by making a three parameter least squares to fit to the speed, cost and time data on the ensemble of computers. This analysis showed that an almost linear relationship between speed and cost was optimal:

$$s = c\ 1.15 \times 1.538\ (t - 59.6)$$

A similar computation was made involving only 29 machines and the power of c was 1.00. The optimal relationship is so close to linear that it will be assumed to be linear. Therefore, the proper relationship between speed and cost is the previous one given under Cost Performance.

Hardware Costs

In order to derive a model for the hardware cost resulting from a given procurement policy, we must define the policy and the workload function of the installation:

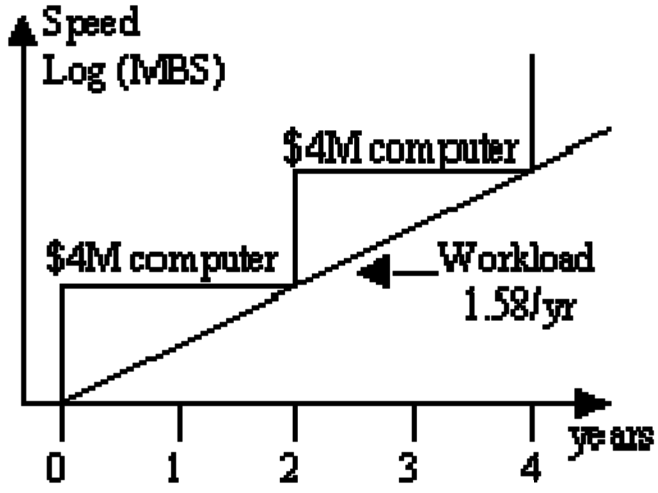
- n = number of years before acquisition of a new computer
- m = number of years the old computer is retained after delivery of the new computer (overlap)
- w = workload increase rate (ratio of workload next year to this)

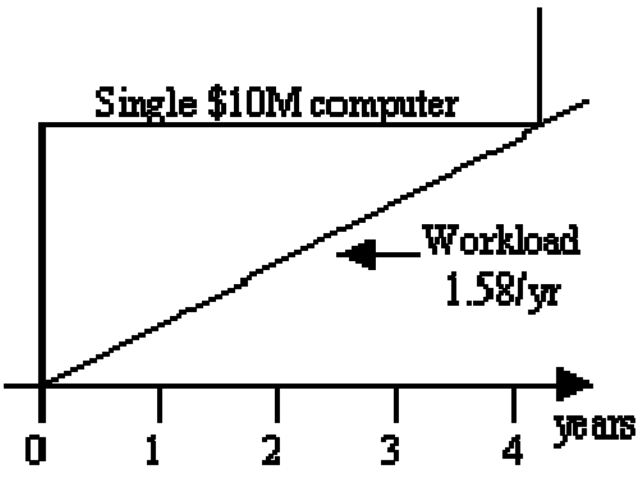
The cost of a computer which must meet the workload n years hence can now be found utilizing the previously derived speed to cost relationship. This results in the following yearly computer costs:

- Purchase: $c = B\ (1/n)wn$
- Lease*: $c = B\ ((n+m)/\ 4n)wn$

B is independent of n, but may vary with time. For a given procurement, B equals the cost of the computer system which would just handle the current workload. If the workload is increasing faster than 1.565 per year, then the installation's computer costs will inflate slowly with time.

$$B = (w/1.565)\ (t - to)$$





The cost functions increase exponentially with the length of the procurement cycle, n. This is because the workload is rising exponentially and the cost per computation is decreasing exponentially. Therefore, considering only hardware costs, it is advantageous to procure new, improved design computers fairly often rather than large computers less often. The charts below show this:

***Based on a four year lease vs. purchase trade-off.**

Software Costs

In order to estimate the effect of different procurement cycles and overlap periods on the software costs, a simple model of software activity needs to be defined. The procurement costs will be broken into two categories; first, the cost associated with conversion to a new machine, including any systems software required, and second, the ongoing programming load of new work. Thus, we have the definition:

A = total software conversion cost to transfer all current programs to a new machine

All of the conversion programming must be completed during the overlap period since it is usually unacceptable to suspend any portion of the work for more than a few weeks. In order for this to be true, the level of programming support during this time must cost at least A/m dollars per year. If m is less than n, this level of support would normally be retained until the next procurement cycle working on the new work. Some of the new work can be done during the overlap period, thereby eliminating some of the pure conversion workload. However, since rethinking the program takes longer than just recording it, the total overlap period workload would be increased, not decreased.

If it is assumed than m and n are fixed, then the programming staff will be forced to grow to at least the A/m level. Alternatively, if it is assumed that the programming staff level is fixed, then the overlap period will be extended to the point where A/m equals the defined level. In either case, the result is that the software costs are inversely proportional to the overlap:

$$\text{Software Cost} = A/m$$

Composite Cost

When the software costs are added to the hardware costs, we obtain a formulization of the composite cost which can be minimized with respect to the cycle time, n, and the overlap time, m. For the least case, this minimization results in very small values of n (less than one year) for all cases where m is less than n. Since the physical constraint of housing three computers simultaneously is considered to be a severe difficulty for most installations, the overlap, m, was limited to at most equal n. This is the full overlap case where each computer is kept until a third computer is delivered. Full overlap is also optimal for the purchase case, since the software cost can be reduced by increasing the overlap, with no increase in hardware costs.

The results of the minimization for full overlap in the lease and purchase cases are too complex algebraically to be very meaningful. However, solutions can be obtained by numerical evaluation and these results are presented in the final graph showing the contour plot of n. From this plot, the optimal value of n can be found given the workload growth rate (w), and the ratio of software conversion cost to hardware purchase cost (A/B). Both purchase and lease results are presented, each in the region where it is most cost effective. If lease or purchase is dictated for some reason other than cost, the appropriate contour lines can be projected into the other region to find the optimal cycle time. For all A/B greater than .5, the graph is for full overlap (m=n), since it is the most desirable policy. For A/B less than .5, the optimal overlap is somewhat less than n:

$$M = n2 / (\mathrm{Ln}\ w - n)$$

If the lease/purchase breakeven point is not four years, the lease results should be interpreted using a value for B which is equal to four years' lease cost. The main effect this has is to move the lease/purchase breakeven point slightly up or down.

Summary of Results

The optimal period for holding a computer is dependent mainly on the workload growth rate and the ratio of software conversion cost to hardware cost. The software-hardware ratio will vary from installation to installation and, to some extent, with the computer selected. In most cases, however, the ratio will be between 0.2 and 3.0. The workload will increase at most installations as fast as it can, limited only by the computer budget. Thus, the range of workload rates should be from 1.565 (constant computer cost) to 1.85 (as fast as field allows – 18% cost inflation per year).

Within these expected bounds on A/B and w, the optimal strategy is to either:

- Lease for 1-1.5 years, procure a new computer and overlap the two for 1-1.5 years. In a procurement evaluation, this means the computers must meet the workload expected in 1-1.5 years and be priced for 2-3 years.

- Purchase a new computer every 2.5 - 3.5 years and hold for twice that long. In this case, the procurement will only need to consider the purchase price and the capability to meet the workload in about 3 years.

The foregoing analysis is for the optimal case, assuming that improved computers are continuously available. Since this is not normally the case, a particular installation may be required to wait an additional year until a good follow-on computer is available. The lease costs are very sensitive to any variance; a delay of over ½ year increases the cost more than 10%. However, the cost is much less sensitive to delays in the purchase case; a delay of 1.5 years is required to raise the cost 10%. Therefore, it is usually safer to purchase, unless a firm follow-on is known.

Another problem with leasing is that the optimal cycle time is so short (1 - 1.5 years) there may not be sufficient programming staff available to accomplish the conversion in that time, even if it would be less expensive. The minimal conversion period can usually be estimated by an installation and this estimate should be used in procurement decisions. If the estimate is more than 1.5 years, purchase is most economic.

Families of compatible computers are a somewhat different case since the technology growth is not as fast and the conversion costs are much lower. The average growth rate of compatible families is only 1.4 per year in speed and 1.3 per year in throughput per dollar. This is because the compatibility requirement rules out many architectural and design changes otherwise possible and improvement must depend largely on faster components. This slower growth rate means that a family which started off on the average cost performance curve would fall outside the two- year zone after 4.5 years. his suggests that families would be discontinued after 4 to 9 years, a theory fully justified by history.

Compatible computers may be utilized to update any computer during its 1-4 year cycle as the primary computer for an installation, as long as the software conversion costs are negligible. Since this lowers the average hardware cost during the cycle, the hardware cost, B, can be lowered proportionately. This increases the ratio A/B and, thus, the optimal cycle period, n, is increased. One example of the effect of this is if compatible machines lower the average hardware cost to ½, thus increasing A/B from 2 to 4, n would be increased from 3 years to 3.5 years.

The results developed herein suggest that an optimal policy for computer procurement be one of a fairly frequent purchase of new computers. If current practice differs from this, it may be due to the continual shortage of experienced programmers, which would increase the minimum conversion time and also dissuade managers from incurring any more conversion costs than absolutely necessary. Thus, non-monetary considerations, such as morale and manpower availability, may tend to stretch the cost-optimal procurement cycle length.

APPENDIX

Future Computers

Computer speed is determined by two factors, circuit speed and memory speed. In both areas, important speedups are currently becoming available. Integrated circuit dual in-line packages are just beginning to appear in computer equipment. These circuits are available in various speeds depending on the logic utilized. Computers being announced this year will mainly be lower logic, such as T2L.

Those intended for 1970-71 are mainly ECL logic, which is 5-10 times faster. Memory speed will also be going up sharply during the next two years due to introduction of semi-conductor memories. These memories will be about the same price as core memories at first, but are 3-10 times faster. Thus, during the early 70’s it is clear that the current rate of improvement in cost-performance will be maintained or exceeded. Thereafter, around 1973, the impact of medium and large-scale integration should accelerate improvement even more.

The advanced computers that should be available during the next three years are as follows:

- STAR-CDC: First delivery to LRL in August 1970; second in January 1971. Pipeline design – can perform up to 50 million 64-bit Adds per second. Logic – ECL integrated circuits. Main memory – 1280 ns core, 512 bit word. Maximum CPU rate – 6400 Megabits/second.

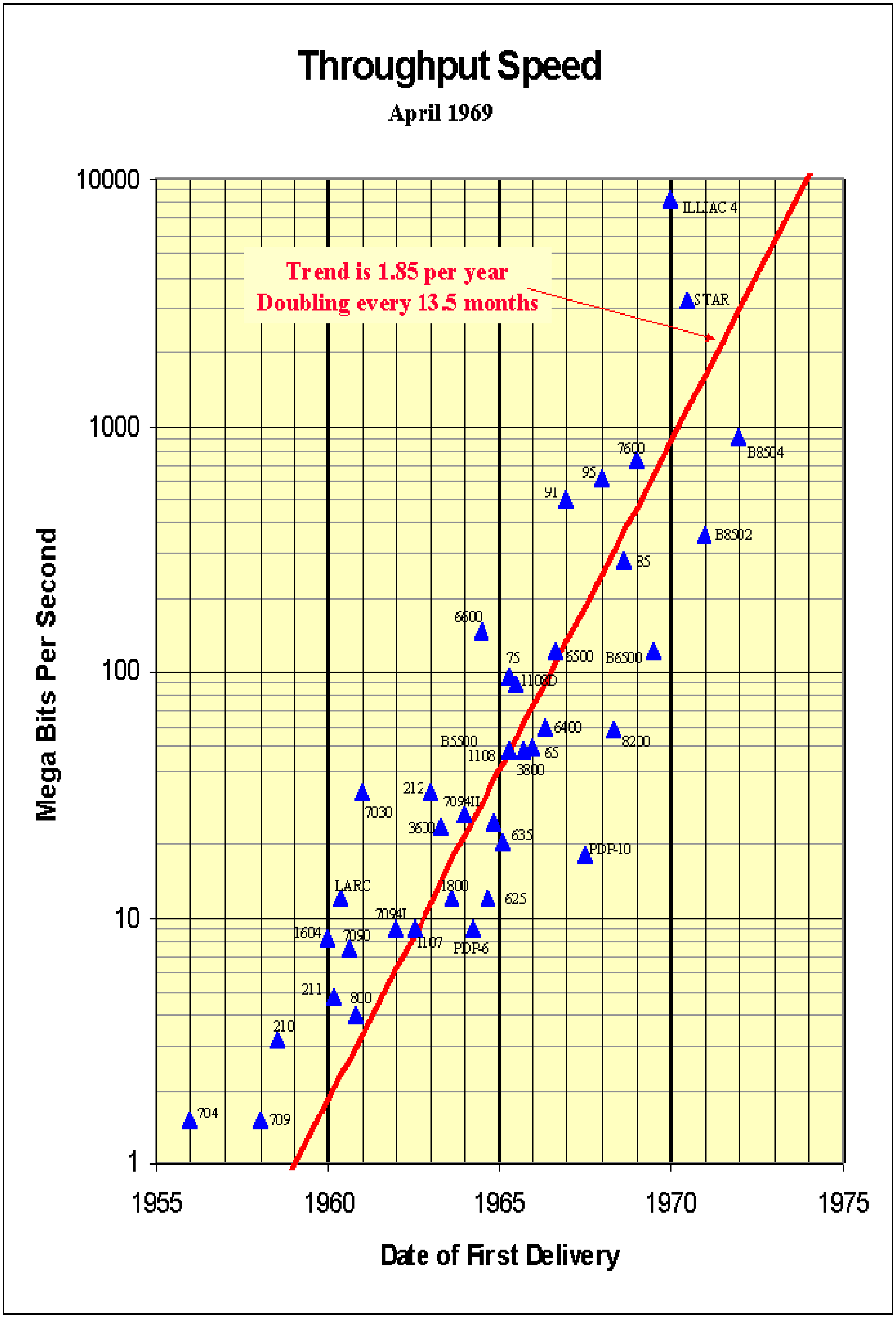
- ILLIAC – Burroughs: First delivery to U of Illinois in July 1970; second in March 1971. Parallel array design with 64 processing elements (PE's) per control unit, up to 4 control units. ECL integrated circuits. Memory 200 ns semi-conductor, 64 bit word, one stack per PE. For 64 PE system, maximum CPU rate – 16400 Megabits/second. Development being sponsored by DoD, thus Government sales will be without large markup.

- ACS - IBM: Very confidential, but presumably 360 compatible and to be announced in early 70's. Speed probably competitive with STAR. Also, three other IBM computers are rumored to be under development.

- ASC – Texas Instruments: Pipeline design, ECL logic, thin film memory. Speed probably comparable to STAR, but details not available. First system operational in late 1970.

- B8500 – Burroughs: Current schedule for the B8502 is January 1971. The B8504 will utilize ECL logic and semi-conductor memories, obtaining a speedup of at least 2.5. It should follow the B8502 within 1-2 years. Both will be compatible with B5500. Designed for parallel operation with many processors. For the B8502, maximum CPU rate = 512 Megabits/second per processor.

As shown above, the equipment due to be introduced in the early 1970's will significantly outperform the current leader (7600) in both raw speed and cost performance. The slippage of some of the developments cannot be predicted, but for the ILLIAC and the STAR, it is clear that slippage can and will be held to less than six months. Thus, within two years at least two computers will be demonstrable which are more than 10 times as cost effective as the 6600 or the 85, and 4-10 times as cost effective as the 95 or the 7600.



Throughput Speed

April 1969

Trend is 1.85 per year
Doubling every 13.5 months

Machine	Installed	MIPS	Wd Lngth	Bits/usec	$ M	B/S/$
CDC						
1604A	1/60	.16	48	7.5	2.3	3.2
3600	6/63	.48	48	23	2.3	10
3800	12/65	1.0	48	48	2.45	20
6400	4/66	1.0	60	60	3.0	20
6500	4/66+	2.0	60	120	4.5	27
6600	9/64	2.4	60	145	5.5	27
7600	2/69	12	60	720	9.4	76
STAR	8/70	25	128	3200	10.0	320
IBM						
704	12/55	.042	36	1.5	1.5	1.0
709 7090	9/58 6/60	.042 .23	36 36	1.5 8.2	1.5 3.1	1.0 2.6
7094I	9/62	.25	36	9	3.4	2.6
7094II	4/64	.71	36	26	3.7	7.0
7030	7/61	.5	64	32	7.7	4.2
65	3/66	.77	64	49	3.4	14
75	11/65	1.38	64	88	4.8	18
85	6/69	4.4	64	280	8.0	35
91	2/67	7.8	64	500	7.7	65
95	2/68	9.7	64	620	8.6	72
UNIVAC						
LARC	5/60	.25	48	12	6.5	1.9
1107	9/62	.25	36	9	2.5	3.6
108	9/65	1.33	36	48	4.0	12
1108 Dual	9/65	2.66	36	96	5.6	17
BURROUGHS						
B5500	11/64	.5	48	24	1.2	20
B6500	6/69	2.5	48	120	2.5	48

B8502	1/71	7.5	48	360	4.3	84
B8504	1/72	19	48	900	5.0	180
ILLIAC 64	7/70	128	64	8200	10.0	820
ILLIAC 256	12/71	512	64	32800	22.0	1500
DEC						
PDP-6	10/64	.25	36	9	1.0	9
PDP-10	9/67	.5	36	18	1.0	18
GE						
625	4/65	.33	36	12	2.4	5
635	5/65	.55	36	20	2.9	7
HONEYWELL						
800	12/60	.084	48	4	1.3	3
1800	11/63	.25	48	12	2.4	5
8200	6/68	1.2	48	58	2.4	24
PHILCO						
210	11/58	.066	48	3.2	1.0	1.7
211	3/60	.10	48	4.8	2.1	2.3
212	2/63	.67	48	32	3.4	9.4

Data Processing Technology Forecast

Dr. Lawrence G. Roberts
Advanced Research Projects Agency