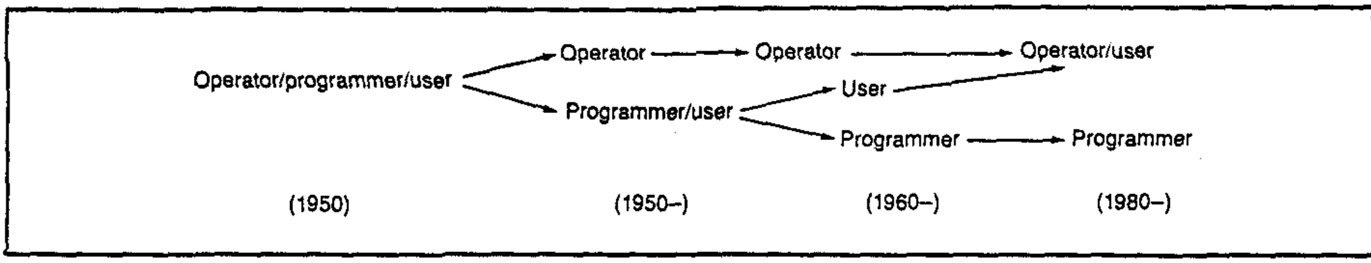
50

The Evolution of Workstations

David L. Nelson and C. Gordon Bell

Introduction

Ironically, the first computers were personal computers dedicated to solving the specific problems of their creator/programmer/user. The notion of a "user" who was distinct from an "operator" did not evolve until machines were sufficiently complex (circa 1950) to warrant this job specialization. Later (circa 1960), the distinction between "programmer" and "user" evolved as applications became sophisticated enough to allow the user of the code to remain unaware of its internal design. Both represent significant steps in the evolution of the human being as he/she adjusted to the environment of advancing computer technology.



So, in one sense, 30 years of technology has allowed us to regain control over the machines we created and then temporarily relinquished to others. We no longer need be at the mercy of operator mismanagement (we have ourselves to blame now), nor the tyranny of the corporate computing centers, nor the bureaucracy of the MIS departments. For we now have our own personal, dedicated computing workstation. It has been a long time, but it was worth the wait.

Some Early Developments

Personal workstations, like other man-made objects, appear strictly evolutionary, going through well-defined stages including the initial idea (1950s); prototypes using large computers to demonstrate the idea (1960s); early but limited use of interactive graphics displays (1970s); working prototypes to test the concept (Xerox Altos—1970s); product introduction by Apollo, Sun, Xerox (1981); full-scale use with too many suppliers (1983); healthy industry beginning a supplier shakeout (1985); steady-state supply to a replacement market (??); decline through replacement of a better idea (??); extinction (??). (A history of personal workstations will be published in [1].)

The context in which this evolution occurs is most complex, with each stage borrowing and rejecting ideas used in other forms of computing. Figure 1 shows the generation of computers since 1950 versus price at introduction. Several tiers can be observed along constant cost regions, and although workstations are shown in a well-defined niche, the concepts that allowed their emergence are deeply rooted as far back as the first generation.

Prior to 1980, there was little agreement over the technical attributes of a personal workstation, and the various prototypes in use had little in common. Then there emerged a consensus of requirements, frequently referred to as the 5M machine, referring to Megabyte memory, Megapixel display, MIP processor, 10+ Megabyte disk, 10 Megabit/sec network. By 1981, machines matching these parameters became

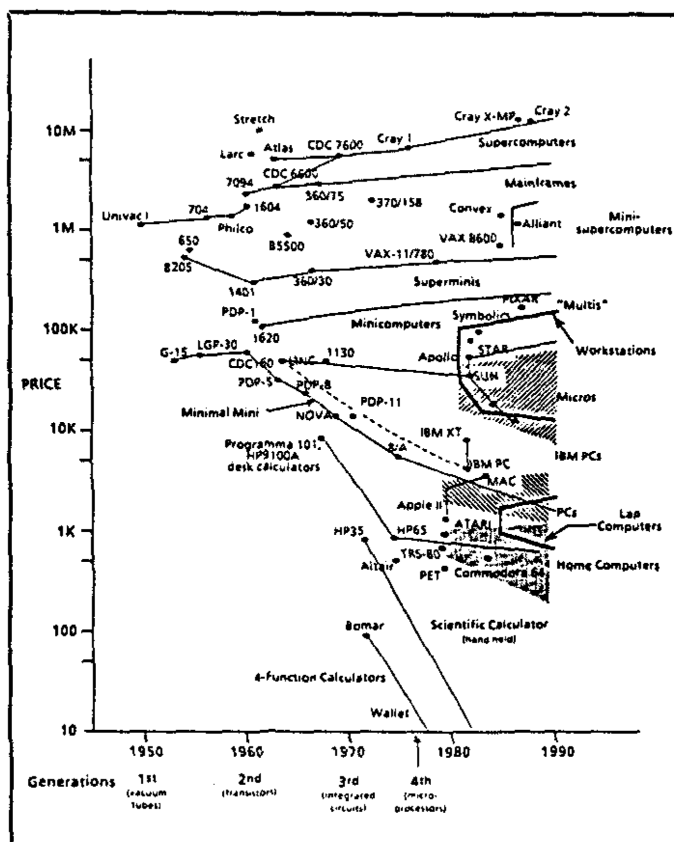


Fig. 1 Computer system prices at introduction and classes versus time.

8755-3996/86/0700-0012$1.00 © 1986 IEEE

commercially available at a price of about $45,000 (Apollo DN100), and since that time, a large number of vendors have been manufacturing similar machines, *lowering the price about 30 percent each year* or adding a corresponding amount of performance or functionality at constant price.

It is significant that in 1980, a common view emerged out of what was previously a disjointed set of ideas and ad-hoc commercial products. What is more significant is that these new machines represented a fundamental change in the way we would use computers—a qualitative (versus merely quantitative) change in our modes of use.

What are some of the forces that determine the timing of these events? We examine a methodology for predicting such changes.

A Methodology for Technology Prediction

In spite of the flurry of technological advances since 1960, most of the parameters describing computers have remained constant, including power densities, cost per chip, cost per pound, etc. For the most part, all of the hardware advances are tributable to the logic densities of silicon and magnetic devices, and, here, the rates of change from year to year have been relatively constant. Taken together, these constants form a simple and accurate method of prediction.

In 1975, the second author used the predicted memory densities to extrapolate system prices versus time for various memory sizes (Fig. 2). That memory size, which is well correlated to functionality and system performance, was well understood, and the model turned out to accurately predict the Apple II (just three years away), the current class of microVAX products, and the new class of minisupercomputers just now emerging.

The first author has used an expanded method for *explaining/predicting technological change and its effects*. This method was actually inspired by the second author more than a decade ago, and is consistent with the three-tier model developed at DEC and now used by IBM. More generally, however, we introduce the tier model through a series of axioms, presumed to be self-evident:

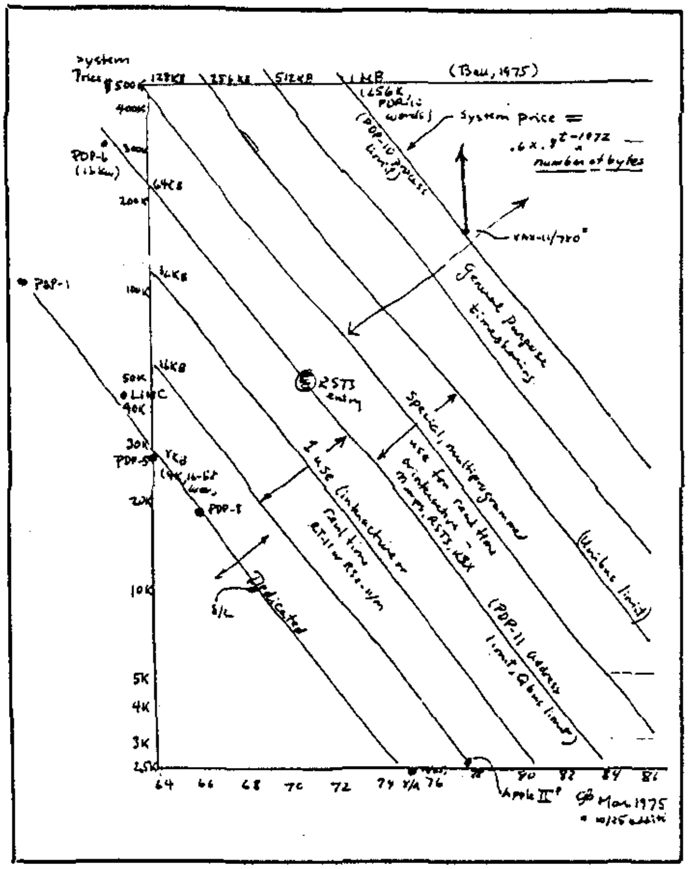


Fig. 2 System price vs. time for various memory sizes and system types (Bell, 1975).

- Axiom 1. Computers cost about $200/lb, independent of their size, valid for the past 30 years when adjusted for inflation.
- Axiom 2. Weight is the principal determinant of use (versus required MIPs, memory, disk, etc.); that is, a "person" uses 50 lb of computer, a "department" uses 500 lb of computer, etc., independent of time.
- Axiom 3. There exists seven tiers (wallet, pocket, briefcase, office, department, center, region), each separated by a factor of ten in weight and, therefore, cost. (See the Table entitled "Nelson/Bell computer classes.")
- Axiom 4. The aggregate technological improvement rate across the tiers is about 35 percent per year, implying that a factor-of-ten improvement

Tier	Location	Name	Price, $	Weight, lb
1	Wallet	Calculator	10	0.05
2	Pocket	Spec function	100	0.5
3	Briefcase	Kneetop	1K	5
4	Office	Workstation	10K	50
5	Department	Shared mini	100K	500
6	Center	Mainframe	1M	5,000
7	Region	Supercomputer	10M	50,000

Table Nelson/Bell computer classes.

(one tier's worth) is achieved every seven years.

- Axiom 5. Tier transitions cause qualitative changes in computing usage, most difficult to predict.

Together, these axioms provide a simple way of correlating (1) technical concepts, (2) modes of use, and (3) time. In particular, it explicitly reveals when a technical concept (such as a hierarchical file system or a megabyte of memory) in one tier will emerge in another. It is a good predictor of when concepts will appear, but it is a poor predictor of how the new tiers will use the concepts. For example, the memory capacities of workstations could have been predicted 20 years ago; however, the need and role of the local area network (LAN) could not have been understood without the experience of time-sharing systems of the '70s. So, while the technology appears relatively easy to predict, new modes of use are difficult to anticipate.

The Workstation Tier — A Fifty-Pound Package

Just as 32-bit mainframes appeared in 1965 (IBM 360), and 32-bit departmental time-sharing systems appeared in 1972 (Interdata 32), so did the 32-bit microprocessor appear in 1979 (Motorola 68000), each event separated by the seven years suggested by their approximate factor-of-ten price separation ($1 million; $100,000; and $10,000 system cost) (Fig. 3). More importantly is the emergence of a virtual memory version at these prices, occurring six years later in each tier: 1970 (IBM 370), 1977 (DEC VAX), 1984 (Apollo DN300).

So, the workstation emerged from the simple downward evolution of well-established systems technology into an office environment (50-lb package). Like most things, however, technology carries a form of baggage with it, and so some questions arise: Is the time-sharing operating system appropriate for a workstation? Should early generation dumb terminals be emulated on the workstations' display? Moreover, how can the user interface reconcile traditional command input with more modern techniques (icons, mice, etc.)? How is work-group interaction — so common in time-sharing — accomplished on workstations?

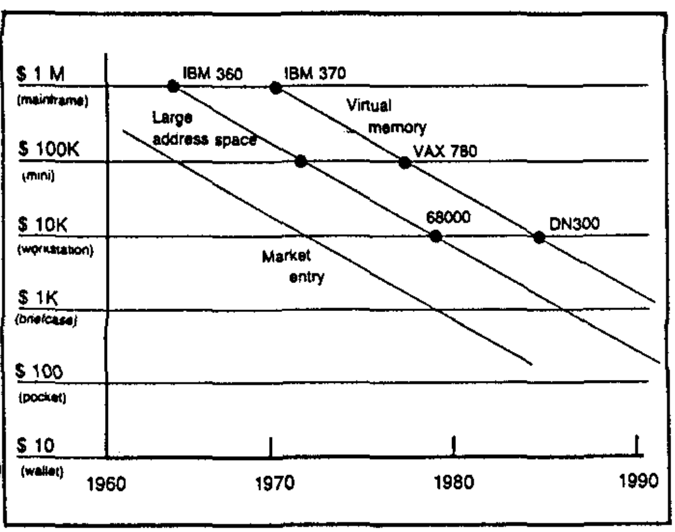


Fig. 3 Technology migration across tiers (×10 every seven years).

A Question of Design Center

As with most man-made objects, computer systems have a *design center* — the point where the system parameters are most balanced and optimum. While the design center of the personal computer is clearly the individual, the design center of time-sharing was the work group (in many ways at the blatant expense of the individual). Can a system have multiple design centers to satisfy both individuals and groups? Workstations, in the end, will clearly have to accommodate both, but the evolution has been mixed.

Viewed as a Personal Computer

Workstations can be thought of as evolving (upward) from the personal computer, best typified by the IBM PC/AT. The machines look similar and it is generally thought that while PCs embrace 16-bit technology and run PC-DOS, workstations embrace 32-bit technology and run Unix™. This view supports the notion of the workstation as a rather dedicated standalone system design to enhance individual productivity. Although networking may be available as an option, it is not part of the system's design center, and, as such, network extensions are awkward and suboptimal.

Viewed as a Replacement for Time-sharing

Workstations can also be thought of as evolving from time-sharing systems. Time-sharing systems rapidly evolved during the 1970s to serve departmental computing where group productivity is as important as individual productivity. When a time-sharing system is replaced by a collection of workstations, the departmental work group runs the risk of losing an important attribute of the system it replaced — work-group computing.

To address this directly, the designers of the Apollo Domain system [2] added yet another axiom — *never take anything good away from a user.* Therefore, if the work group had grown accustomed to time-sharing, it is important to preserve its valued attributes and, hence, the adoption of the distributed file system as an integral component of the design. More generally, the designers developed a dualistic design center — first, the workstation should be design-centered to maximize individual productivity; second, the collection of workstations should be design-centered to maximize group productivity. The Apollo Domain used the single-level-store (SLS) model [3] to implement the file system because it offered maximal transparency for both code and data, and, to the maximum extent possible, it preserves the file-system semantics of a typical time-sharing system, even solving the complex problems of file locking and concurrency control.

The preservation of time-sharing file-system semantics in a work group of workstation users necessarily requires a local area network. Here, the purpose of the LAN is beyond that of simple communication found in conventional networks. LANs of workstations require a higher degree of integration, including distributed file access, remote procedure calls, cooperative computing, etc. The functions of these capabilities are inherent in time-sharing and have to be recreated for the distributed environment—a technical task that is industry-wide and far from completion.

The Adoption of Workstation Standards

Since 1981, workstations have developed around a set of standards which has differentiated them from lower cost 16-bit PCs. The most important standards are (1) 32-bit architecture, (2) Unix operating system, and (3) Ethernet local area network. While these standards offer a base-level capability and serve to distinguish workstations from PCs, in many ways they are limiting the broader acceptance of the systems. Many, if not most, applications do not require 32-bit architectures, and the complexity of UNIX compromises ease of use and adoption by non-computer professionals. The dichotomy of PCs (16-bit PC-DOS) and workstations (32-bit Unix) is a technological artifact, one that is becoming more and more difficult to sustain. We can look forward, therefore, to the merging of 16- and 32-bit systems.

The Merging of Graphic Terminals, Workstations, and PCs

Prior to 1986, workstations, in general, had several subclasses that were largely distinguished by application: (1) low-end 16-bit PCs for spreadsheet and word processing, (2) 5M workstations for computer-aided-engineering software development, and (3) high-performance graphic terminals for three-dimensional computer-aided-design applications.

Two trends are occurring to break down these classes: first, the technology required to implement 5M-class workstations is rapidly becoming available in the 16-bit PC price range, and, second, the maturing high-end graphic technology, previously embodied in dedicated terminals, is being integrated into conventional 5M workstations. Consequently, a vertical expansion of the 5M-class machine is occurring (Fig. 4), placing downward pressure on the 16-bit PCs and upward pressure on the high-end graphic terminals, clearly suggesting the demise of both.

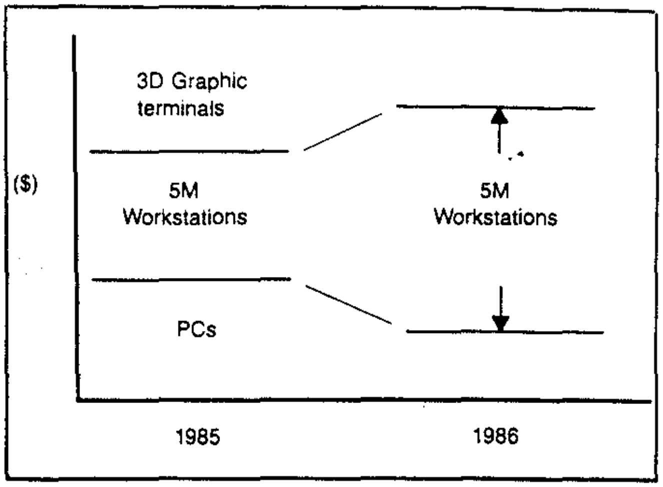


Fig. 4 Vertical expansion of 5M-class workstation.

The LAN—The System Backplane

The pioneering work done at Xerox Palo Alto Research Center laid the groundwork for viewing the LAN as the systems bus. The prototypes of personal workstations (Altos) were interconnected with an early version of the Ethernet and complimented with network-based servers for computing, file storage, printing, etc.

More than ten years later, the computer industry is embracing this model through the alliance and standardization of complimentary components:

—the LAN itself, destined to be Ethernet, a token ring, or a broadband network—combinations of all, working together;
—workstations of various types, but running applications on standard environments (UNIX, PC-DOS, LISP);
—computational servers (minisupercomputers) providing high-end computing to the workstations;
—file servers for bulk storage and backup;
—communication servers, bridges, and gateways to handle geographic spreads, access to older machines, etc.;
—servers connected to laboratory and factory facilities that will help automate complex processes.

Conclusion

In summary, the workstation is merely a component in a much more complex and evolving system. For all of this to be successful, the computer industry and its users must reconcile the many problems of interconnecting heterogeneous components through standard protocols and interfaces.

Some have estimated that workstations have penetrated a mere 2 percent of their potential in terms of number of units. Applications are still at formative stages, particularly with respect to the human interface and distributed computing. Finally, major advances in the utilization of artificial intelligence techniques will require growing amounts of workstation capabilities.