HEWLETT-PACKARD
JOURNAL

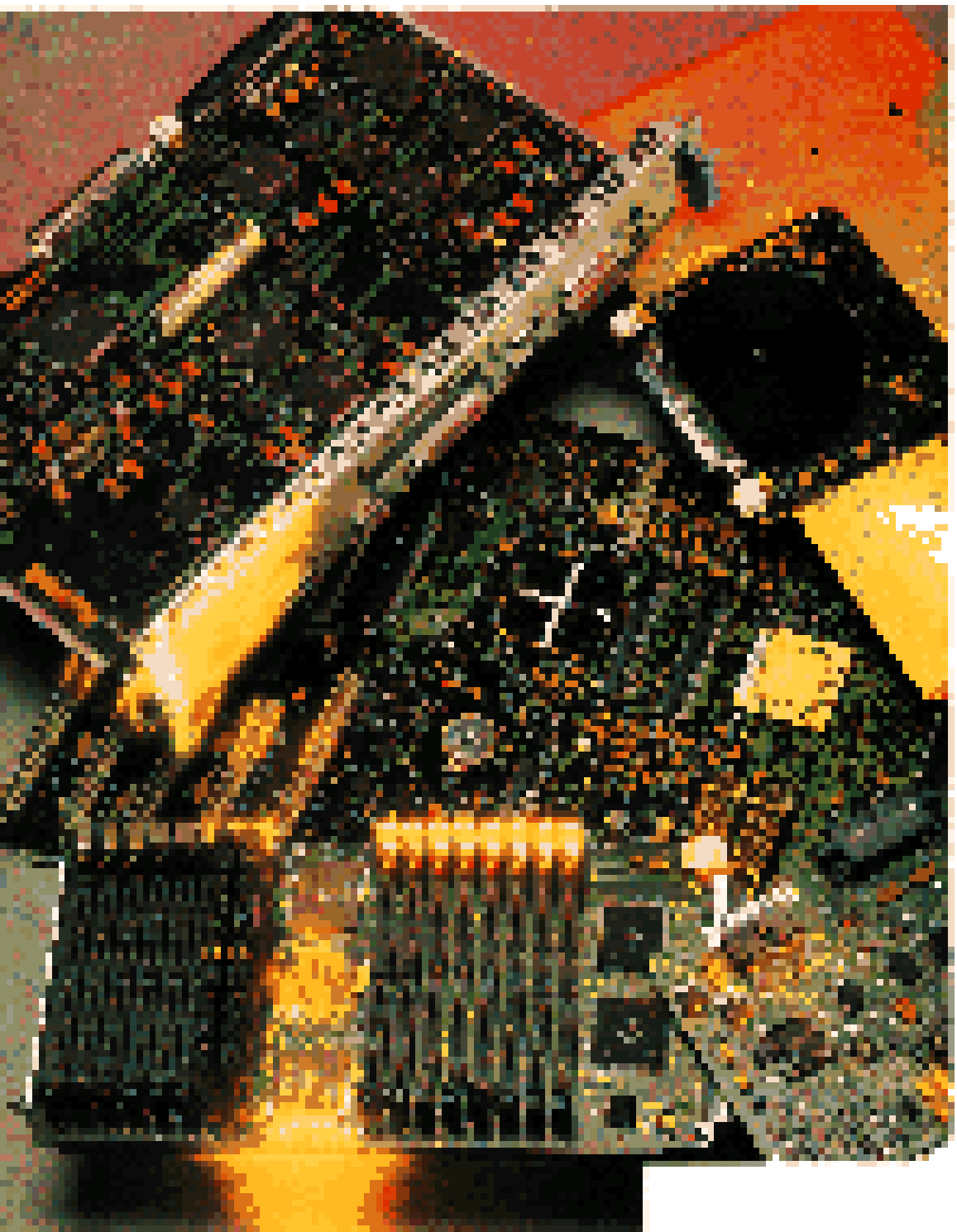


table of contents

June 1995,
Volume 46, Issue 3

Articles

Capillary Electrophoresis: A Product of Technological Fusion

An introduction to capillary electrophoresis (CE), its different forms, and its applications, and the history of CE research at HP, leading to the new HP CE instrument described in this issue.

by Robert R. Holloway

Electrophoresis is one of the most powerful, if traditional, analytical methods. Picton and Linder, Hardy, and Ellis were turn of the century pioneers in the analysis of biocolloids (proteins and carbohydrates). By electrophoretic methods they were able to work with these previously intractable materials. Arne Tiselius ushered in the age of instrumentation with a demonstration of the first cell for electrophoretic analysis in the 1940s, and in the 1990s electrophoresis is one of the most visible icons of science, with TV and newspapers daily displaying electrophoretic DNA spot patterns.

The fused silica capillary, a glass tube about the size of a hair, is a spin-off of the optical fiber. It exists more because it could be made (by heat-drawing a large glass tube into a tiny one) than because it was seen as a powerful analytical tool. It is not surprising that one of its grandest applications, capillary electrophoresis (CE), was not foreshadowed.

Various workers in the electrophoretic field (Everaerts, Hjerten, Mikkers, Virtanen), while aware of the benefits of going to small systems, did not employ the fused silica capillary. The growth of the method began when a few workers, including scholars Jorgensen and Lukacs of the University of North Carolina, industrial researchers McManigill and Lauer of HP Laboratories, and others, not necessarily sophisticated in electrophoresis but aware of the power of the capillary in gas chromatography (demonstrated by Ray Dandenau and Ernie Zerenner of Hewlett-Packard in 1979[1]), conducted the first experiments in fused silica.

Separation Science

In analytical chemistry, separation is a fundamental process. A chemical substance is generally intimately mixed with many other substances, and its determination and identification is made considerably easier by its physical separation from the mixture. Thus, analytical chemists have worked hard to understand the separation process and to develop many modes of separation.

The separation principle for a particular separation method is the physical or chemical property that varies in magnitude among the substances in a sample. In chromatography, for example, the separation principle is often chemical affinity for chromatographic materials. A feature of CE is that its separation principle is orthogonal to that of liquid chromatography (LC), or in other words, it has a completely different basis.

Separation implies physical movement or transport. In free solution capillary electrophoresis (FSCE), the simplest form of CE, transport of a particular chemical species is the resultant of several driving forces. The species moves in response to an electric field, which is important if it is charged. It moves in response to the flow in the channel, which can be caused by the electric potential difference across the fluid-silica interface (electroosmotic flow) or by mechanical pumping of the fluid. Finally, its migration is affected by the frictional drag it experiences, which depends on its size and shape and the viscosity of the fluid.

Modes of CE

CE is really a group of several procedures. CGE, or capillary gel electrophoresis, is an enhancement of an older technique. IEF, or isoelectric focusing, has also been done in other formats, and is still in the process of adaptation to the capillary. ITP, or isotachophoresis, has not found wide use but has considerable potential for preparing pure chemicals. MEKC, or micellar electrokinetic chromatography (also called by other acronyms), is a completely new method that combines electrophoresis and partition chromatography. The simplest and most characteristic mode, in terms of which all the others can be described, is FSCE. It is the form most practiced today.

Free Solution CE

In FSCE, a capillary channel, typically a few micrometers to two hundred micrometers in diameter, is filled with a conducting liquid, most often a water solution containing an acid or a base and its salt, termed a buffer, which has a pH that is insensitive to the addition of small amounts of acids or

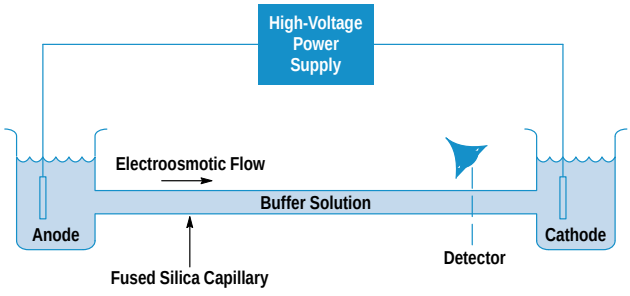


Fig. 1. Free solution capillary electrophoresis.

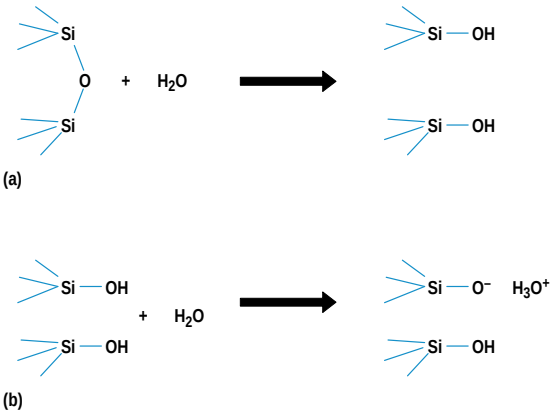


Fig. 2. (a) Hydration of a raw silica surface to polysilicic acid. (b) Negative charging of silica by ionization of the surface.

bases. The two ends of the channel are immersed in two reservoirs which are held at different electric potentials; in other words, a voltage is applied to the column. In the most usual situation, this will result in the flow of the liquid from the anodic reservoir to the cathodic reservoir. This is electroosmotic flow (Fig. 1).

Electroosmotic Flow

The surface of silica in contact with an aqueous medium with a pH no lower than 2.5 or so is loaded with negative charge. This is because silica hydrates and becomes an acid, releasing positive hydrogen ions into the medium (Fig. 2). The excess positive charge is physically localized within a very narrow zone close to the surface (Fig. 3).

Since capillaries enclose very small channels, and since electrophoretic currents are proportional to the cross-sectional area of the channel, FSCE involves small currents (from a fraction of a microampere to a hundred microamperes or so) and thus small amounts of heat relative to conventional scales of electrophoresis. As a result it is possible to apply much larger voltages, and axial electric fields in the channel are high—hundreds of volts per centimeter.

Water molecules in the narrow zone of positive charge are subjected to the pull of the positive charge as it moves to the cathode. In fact, the entire column of fluid is dragged. In contrast to what would happen if this same column of fluid were moved by a piston, there is no laminar flow, no wall drag tending to produce a parabolic flow profile (Poiseulle flow). The flow is proportional to the electric field, and can reach a velocity of millimeters per second.

Transport

At the anode a thin zone of a mixture of analytes in aqueous solution is introduced into the capillary. The molecules in the zone are immediately subject to the strong electric field and the bulk flow of the electrolyte. They are impeded by the molecules of electrolyte through which they must move.

The different species each move with a characteristic velocity. In general, they move toward the cathode, but if strongly negative, can actually be expelled into the anodic reservoir. As they move, they separate into bands enriched in the various species relative to the original mixture band.

The most spectacular feature of CE is that the molecules in a band are subject to very little dispersion while traveling along the column. The flow in the column does not appreciably stir up the band, because it is a flat-profile flow. Diffusion is inescapable and does spread the band, but because the experiment is very fast, not much diffusion occurs. In the CGE form of the technique, the medium is a gel, diffusion is drastically slowed, and prodigious resolution is achievable.

Somewhere along the tube, as close as possible to the cathode, a detector is mounted to visualize the bands as they pass. Since silica is transparent to ultraviolet radiation, the most convenient detector consists of a lamp on one side of the column and a photodetector on the other side. The bands absorb the light as they move into the path of the lamp rays, and the signal is displayed as a series of peaks similar to the familiar peaks of chromatography, although unlike chromatographic peaks the quantity of material represented by a particular peak depends on when it passes the detector as well as its size. Figs. 4 and 5 show typical electropherograms obtained by FSCE and CGE, respectively.

CE Applications

The mature practice of electrophoresis and the new technology of fused silica capillaries have been combined to perform previously unrealized analyses (CGE for the quantitation of oligonucleotide failure sequences and automatable DNA sequencing), to perform others with unprecedented speed and resolution (faster DNA restriction fragment length poymorphism (RFLP) analysis and peptide mapping than liquid chromatography), and to provide a badly needed orthogonal separation principle for all liquid-phase analysis (the combination of charge and frictional drag). CE is not just an incremental improvement on existing methods; rather, it allows entirely new things to be done. It is this paradigm-shifting character that makes CE attractive in an industry

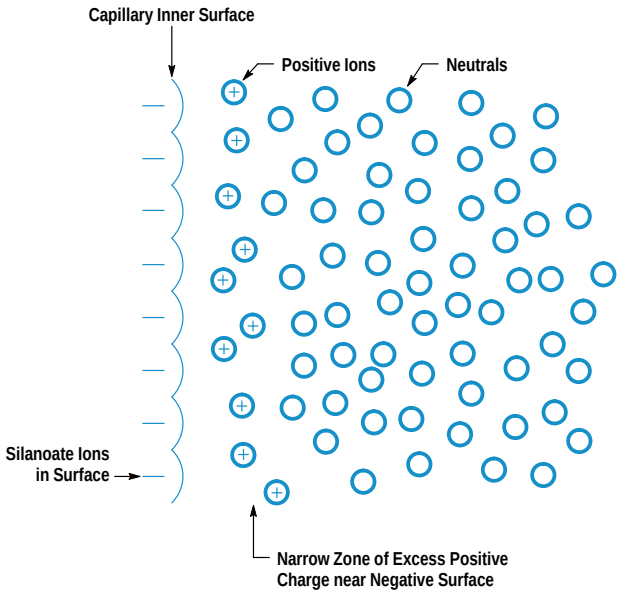


Fig. 3. In electroosmotic flow, a double layer of positive and negative charges forms near the surface of the capillary.

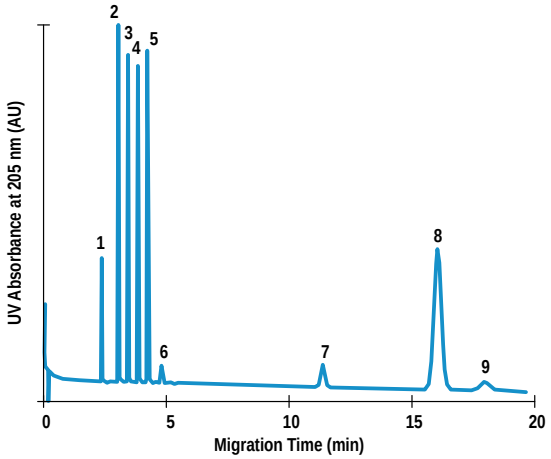


Fig. 4. An FSCE separation of an amine mixture. Peaks: 1. pyridine. 2. benzylamine. 3. diphenylamine. 4. adenine. 5. aminophenylboronic acid hemisulfate. 6. 4-aminoantipyrine. 7. xanthene. 8. 2-amino-5-nitrophenol. 9. hypoxanthine. Buffer: 0.1M phosphate, pH 2.7. Separation conditions: anode 20 kV, cathode 10 kV. Column: 40 cm. Injector-detector: 15 cm.

that, while not growing as explosively as the computer industry, has been changing rapidly, with products becoming obsolete in a matter of months.

Most of the separations performed in the mainstream chemical industry, as well as the food and drug industries, can be carried out and are being tried by CE. In some instances, notably the large-molecule areas of bioscience, good applications have already been found and are on their way to validation. In many others, sometimes surprising ones involving molecules with no charge differences, there is great interest. The high resolving power of CE sometimes outweighs an inappropriate separation principle, for example in the case of chiral drug molecules. These molecules are uncharged and have the same frictional drag and therefore cannot be separated with ordinary media. Their separation in CE comes from transport through a medium that is itself chiral, and the low dispersion of electroosmotic transport preserves it for examination.

CE and the Bioanalytical Market

In the biotechnical segment of the analytical chemical market HP has had only a few products. One of the most important is the liquid chromatography (LC) system, which is used to characterize chemical products and validate their identities (traditional quality control). LC is very powerful in the field of general chemicals, along with gas chromatography (GC), but the new bioengineered pharmaceuticals, the nucleic acids, and peptides and proteins challenge even its power. CE is a welcome partner method that shows promise in solving some of these challenges.

Oligonucleotide Failure Sequences

The oligonucleotide, which is a sequence of molecular letters of DNA code, is used in the research and development effort in the new biotechnical industry. These segments, a few to perhaps a hundred letters in length, are synthesized to be used as templates, or more properly stencils, for operations that create large amounts of the informational molecule (polymerase chain reactions or PCR) or as implements for the manipulation and recognition of critically important DNA molecules, such as those that cause inherited traits (genetic diseases) or those that can identify individuals. The synthesis of these molecules is carried out one letter at a time and is prone to mistakes at each step. After a series of steps, a population of failure sequences exists along with the desired materials. These mistakes are most easily detected and quantitated by capillary gel electrophoresis (CGE). Neither the traditional slab gel techniques of electrophoresis nor LC has the resolving power to distinguish the authentic material from these attendant failures.

Peptide Mapping

In FSCE, the most important mode in terms of the number of analyses carried out, an important bioanalytical target is the protein digest. Most of the bioengineered pharmaceuticals being produced or contemplated are proteinaceous. The identity of a protein product is determined by using specific proteins to break it down catalytically into a mixture of smaller proteins (peptides) which identifies its origin just as a geometrical pattern of closely spaced curved ridges identifies the fingertip on which it occurs. The peptide mixture can be analyzed with LC to produce the molecular analog of the fingerprint, called a peptide map. The mixture can also be analyzed with CE to produce an equally characteristic map more quickly. Both maps are useful, and at this stage of manufacturing practice, their complementarity is beginning to be widely appreciated.

The Future of CE: Integrated Liquid-Phase Analysis

CE will continue to find applications. Its growth rate as a liquid-phase analytical method will be higher than its more mature cousin, LC. Most of LC's applications are firmly entrenched and sufficient for immediate needs. Improvements in these will be incremental enhancements involving a variety of techniques including CE.

CE in the future will likely be a part of a complex strategy. Workers are now combining the unique separating power and efficiency of CE with the versatility of LC and the mass spectrometer's ability to identify unknowns. New forms of separation continue to spin off the CE experiment (most recently CEC, or capillary electrochromatography) and will also play a part in the integrated future of liquid-phase analysis.

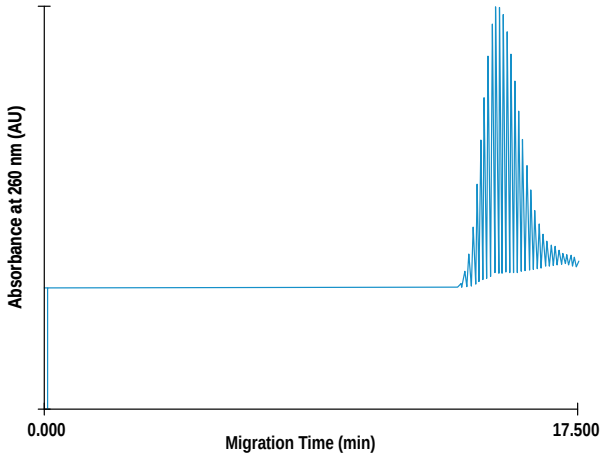


Fig. 5. A CGE separation of oligoadenylic acids. The oligomers range from 40 to 60 units in length.

HP Laboratories and CE

Douglass McManigill of HP Laboratories became interested in CE as a result of the publication of Professor Jim Jorgenson's first paper describing the use of fused silica capillaries for electrophoresis.[2] McManigill, in what is now the chemical systems department of the Analytical/Medical Laboratory of HP Laboratories, was working in the area of supercritical fluid chromatography, but was intensely interested in the physics of transport processes. Along with Henk Lauer of the HP Waldbronn Analytical Division, he built a CE apparatus and began an investigation of the CE separation process.

At roughly the same time, Paul Bente and Joel Myerson of HP Genenchem were working on an electrophoresis instrument to do gel separations of proteins and peptides. While they had limited success instrumentally, the pressurized gel synthesis they invented was the only reproducible technique for making the gels for several years.

In time, Lauer left the company, and over several years, McManigill led the project, working with many others, including Sally Swedberg, who later worked in Waldbronn on CE, mechanical engineer Stu Lerner, who is no longer with HP, Jim Young, mechanical engineer and designer whose contributions included a method of coating columns with a resistive layer, Mark Bateman, computer scientist and electrical engineer, who created the first software for the instrument, mechanical engineer John Christianson, who worked on the first prototype and is now at the HP Vancouver Division, the author, a chemist from the medical group at HP Laboratories, Don Rose from Jorgenson's group at the University of North Carolina, Jürgen Lux from Max-Planck Institut für Kohleforschung in Mülheim, Germany, Cathy Keely, an engineer who made hundreds of CE runs to explore the unexpected behavior of columns with a conductive coating, Tom van de Goor, a former student of Everaerts in Eindhoven, the Netherlands, who also joined the study of the control of electroosmotic flow, and Wes Cole, an electrical engineer who modeled electric fields at the ends of capillaries.

Another group at HP Laboratories, led by Gary Gordon, worked on several issues surrounding a CE instrument. Physicist Dick Lacey helped design a detector. Gary developed the "bubble cell," which offers a greater optical path length. An automated system for the fabrication of bubble cell capillaries was developed by Rich Tella, Henrique Martins, Bill Gong, and Frank Lucia of the manufacturing systems and technology department.

Christmas season 1989 saw the transfer of HP Laboratories technology into the capable hands of Alfred Maute, Fred Strohmeier, Martin Bäuerle, Fritz Bek, Franz Bertsch, Bernhard Dehmer, Monika Dittmann, Ulrike Jegle, Patrick Kaltenbach, Alwin Ritzmann, Werner Schneider, Klaus Witt, and Hans-Peter Zimmerman of the Waldbronn Analytical Division. The result is the instrument described in this issue.

The help of Doug McManigill is acknowledged in the construction of this history.

References

1. R. Dandenau and E.H. Zerenner, "An Investigation of Glasses in Capillary Gas Chromatography," *Journal of High–Resolution Chromatography and Chromatography Communications*, Vol. 2, 1979, p. 351.

2. J.W. Jorgenson and K.D. Lukacs, "Zone electrophoresis in open–tubular glass capillaries," *Analytical Chemistry*, Vol. 53, 1981, p. 1298.

A New High-Performance Capillary Electrophoresis Instrument

This instrument automates the CE separation process with high reproducibility of analytical results such as peak areas and migration times. A diode array detector with an optimized optical path including a new extended lightpath capillary provides spectral information with high detection sensitivity. The liquid handling and sample injection systems are designed for flexibility and usability.

by Fred Strohmeier

Electrophoresis on planar gels (gel electrophoresis) has been well-known for decades as a scheme for separating sample constituents. Capillary electrophoresis (CE), on the other hand, is an entirely new type of separation methodology. The challenges for instrument design stem from the fact that the separation takes place in a capillary with an inner diameter of approximately 25 to 100 micrometers. Samples analyzed in such instruments have a volume of a few nanoliters, while the total volume of the capillary is a few microliters. Miniature dimensions like this require totally different ways of thinking in terms of technical implementation.

Project and Product Goals

When HP made the strategic decision to invest in system development for the electrophoresis market (instrument, workstation, and capillaries, including the related chemistry), it was evident that CE as a separation methodology was still in its infantile stage. It was either just being investigated or, in advanced instances, being used for new application development, for example in the search for new pharmaceutical substances. Most instruments were homemade systems composed of barely modified HPLC (high-performance liquid chromatography) equipment with an additional high-voltage power supply. Analyses were performed in an almost manual manner. As applications were developed it became obvious that instrumental imperfections were beginning to limit further acceptance of CE. Many complaints and shortcomings were stated by potential and real CE customers, who most often worked in an environment where HPLC was the method of choice for most analytical tasks. The most important limitations were missing functionality, lack of automation, not enough sensitivity, and low performance.

There were a few commercial instruments at the time, each type designed to solve a specific CE problem, but no attempt had yet been made to address all of the perceived apparatus limitations in one instrument. Before the HP instrument development was started the user needs were confirmed through numerous customer visits in different market segments. For the project team it became apparent that we had to make significant improvements in the state of the art in the following areas:

- Automation of the overall analysis process
- Detection sensitivity (ultraviolet/visible light absorption)
- Compound identification through spectral information
- Reproducibility of the analytical results (migration time, peak area)
- Usability (mainly capillary handling).

All development activities were driven by these goals. For instrumental functions that were not directly related to this priority list we tried to find an off-the-shelf solution within or outside HP.

Instrument Architecture

To provide total control over all parameters relevant to the quality of a separation, an integrated architecture was chosen. All of the functional modules such as the detector, autosampler, injection module, and capillary are within one mainframe (see Fig. l) where all of the interfaces are well-defined. This architectural approach makes it much easier to implement a single point of control concept. However, the internal structure of the instrument is modular as shown in Fig. 2.

The instrument is basically divided into two major modules: the detector module and the separation unit. They are kept strictly separate not only from a hardware point of view but also with respect to control (each has its own HP-IB interface). This structure provides the flexibility for future upgrades such as adding other detectors to the system. The coordination of the analysis, in which both modules have to be synchronized, is handled by the PC-based HP ChemStation. From there the method set up by the user is downloaded to the detector and the separation unit, and when the analysis is started the method is executed simultaneously by both modules. There is no communication between these two modules except for external start/stop, not ready, and error signals, which conforms to an HP standard for intermodule communication.

A physical connection between the modules is provided by the capillary. The inlet end of the capillary is adjacent to the separation unit, allowing the autosampler to immerse the capillary tip into the sample or a buffer. The major part of the capillary is within the cassette for thermal control purposes. Before it leaves the cassette again the capillary is guided through the optical path of the detector, where the

Fig. 1. The HP G1600A CE instrument is contained within a single mainframe and uses a PC-based HP ChemStation for control.

sample undergoes detection. The outlet end, like the inlet end, is adjacent to the separation module where the autosampler has access to the capillary tips.

Separation Environment

From a usability perspective the most critical design aspect is the way the capillary is interfaced to the infrastructure required to perform an electrophoretic separation (see Fig. 3).

Besides filling the autosampler with sample and buffer vials, the most frequent user interaction with the instrument is the changing of the capillary, either because the application has changed or because the lifetime of the capillary has expired. Before insertion into the instrument, the capillary is positioned in a forced-air-cooled cassette. This design approach has many benefits for the user. Using air as a cooling medium makes the system totally uncritical compared to a liquid cooling system whose connectors are prone to leaks. Even though liquid cooling in principle has a higher cooling efficiency, this benefit is outweighed by the fact that only highly electrically resistant cooling liquids (e.g., perfluorcarbon) can be used and these have a lower thermal conductivity than water and are very expensive. With air cooling the cassette design can be kept very simple because there are no special connectors to interface to the cooling fluid other than conduits to allow the airstream to flow through. Thus, cassettes can be exchanged without any precautions. All critical interface requirements, such as sealing the capillary against the injection vial and aligning the capillary with respect to the optical path, are shifted to the instrument where they are easier to solve. For all of these reasons the price of a cassette is very reasonable. The cassette is designed so that the detector interface is self-aligning with respect to the optical axis. The capillary ends are caught and guided into final position when the user slides the cassette into the instrument. Sealing against the fluid system is automatic.

For high-efficiency cooling, the airstream temperature is controlled using Peltier elements. The forced air allows fast equilibration of the system in case the temperature setpoint changes. The operating range is from 10°C below ambient temperature to 60°C. The temperature is sensed within the airstream. The heat exchanger temperature is sensed to measure cooling efficiency, which is optimized by adaptive control algorithms. The overall precision of the temperature control system is ±0.1°C. Because of this precision, the repeatability of migration time and peak area measurements is very good.

Detection

The detector built into the HP CE instrument is based on a UV/Vis (ultraviolet and visible light) absorption detection scheme. A diode array spectrograph provides parallel readout of all wavelengths shining through the capillary, thus

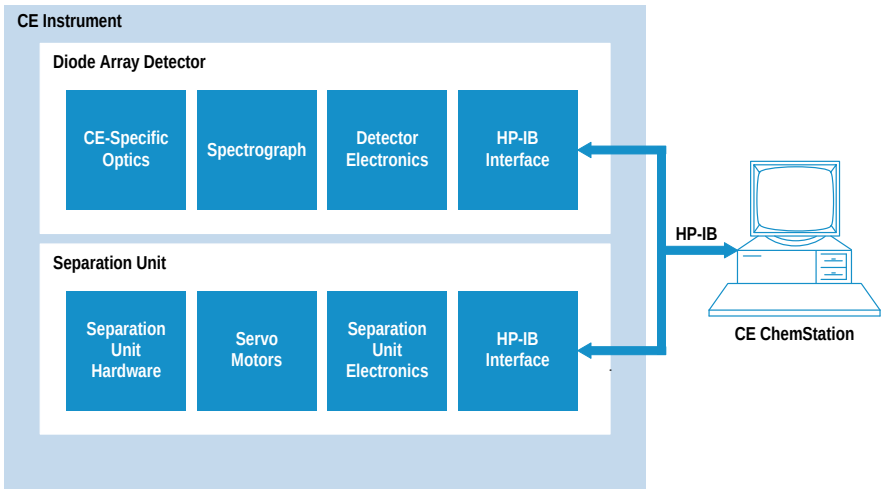


Fig. 2. HP CE instrument architecture.

Capillary Electrophoresis Applications

Capillary electrophoresis was initially regarded as an analytical separation tool for proteins and peptides. Its characteristics imply that biomacromolecules theoretically should take the biggest advantage of the technique. However, it has turned out that more then a decade after the birth of the technique the applications have spread into many more areas than just the bioscience area. In fact, for many proteins it has proven to be a bit of a problem to get a separation with the required high sensitivity using the fused silica columns. All in all this has not hindered the growth of the technique. When the first commercial instruments became available in 1990 the market was estimated to be several million dollars in size. In 1994 the expected market size might very well reach 50 million dollars. The main user groups of CE are found in the pharmaceutical market (both traditional and biopharmaceuticals), the bioscience market, and the chemical industry (see Table I).

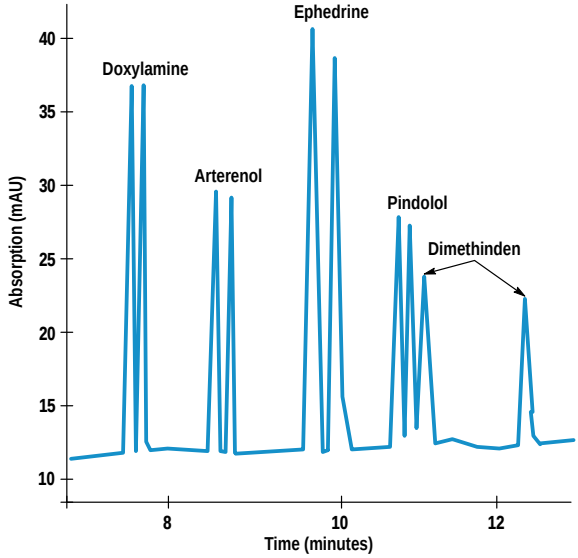
Although still mainly in use in R&D laboratories, the technique is definitely migrating towards controlled analytical laboratories such as QA/QC and product testing labs. This indicates that the technique does offer unique benefits and can expect sustained growth in the future.

Table I
CE Users

Market Segment	Estimated Share (%)
Pharmaceutical Industry	35
Bioscience	35
Chemical Industry	20
Food/Beverages	5
Others	5

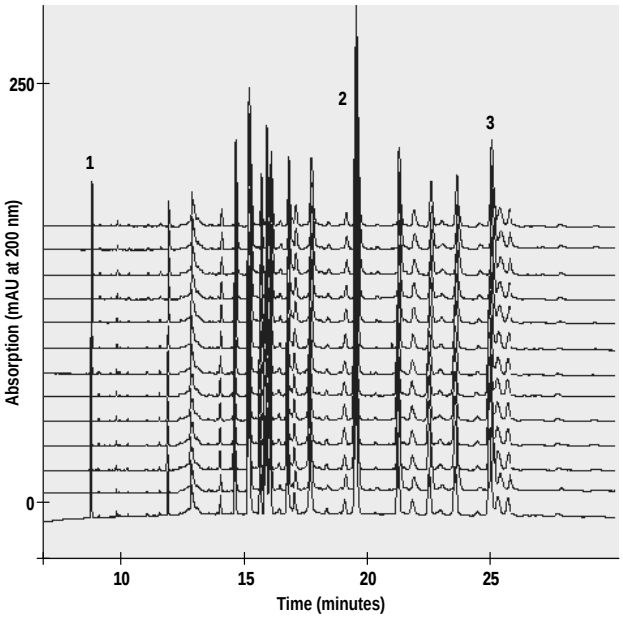
Some successful applications of CE include:

- Analysis of optical impurities (chiral analysis) (see Fig. 1)
- Tryptic digest analysis of recombinant biopharmaceutical drugs (peptide mapping) (see Fig. 2)
- DNA analysis (e.g., PCR product analysis) (see Fig. 3)
- Organic acid analysis (e.g., in beverages) (see Fig. 4).



Sample: Chiral Mixture
Buffer: 20 mM citrate, pH 2.5, 2% Carboxymethyl-β-CD
Capillary: L_{eff} = 56 cm, L = 64.5 cm, i.d. = 75 µm
Injection: 200 mbar · s
Electric Field: 300 V/cm
Detection: Signal 214.20 nm, Ref. 450.80 nm
Temperature: Capillary 20°C

Fig. 1. Capillary electrophoresis (CE) separation of a mixture of basic chiral drugs using cyclodextrin as chiral selector.



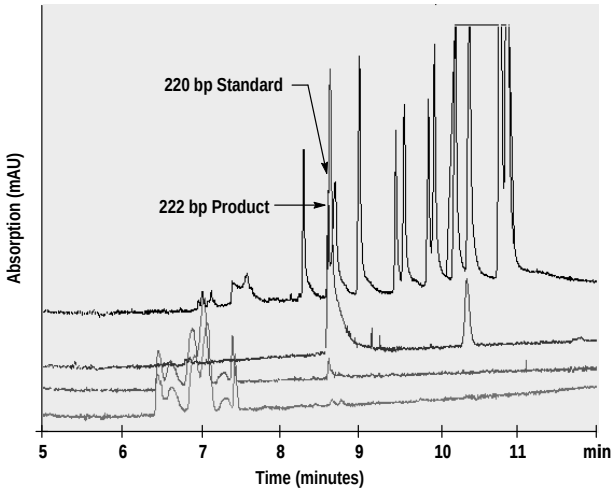
Reproducibility (%RSD)

	Peak 1	Peak 2	Peak 3
Migration Time	0.36%	0.60%	0.33%
Area	1.63%	1.89%	2.09%*

Fig. 2. Repetitive separation of a tryptic digest of recombinant human growth hormone by CE.

providing spectral information. This type of detector has several inherent features that have proved useful for many applications:

- Identification. Absorption spectra make it possible to positively identify substances by their spectral "fingerprints." The HP ChemStation controlling the instrument has a built-in spectral library. Library searches can be performed, resulting in suggestions of substances that have similar spectra. They are ranked according to a computed match factor.
- Confirmation. Spectra created by liquid chromatography separations and capillary electrophoresis separations for the same substance are, with a few exceptions, identical. Based on this, the spectra obtained with these two separation techniques are confirmatory or redundant to each other for a given sample constituent. Since the two separation techniques have different separation mechanisms it is very unlikely that a sample constituent or impurity will be missed by both LC and CE.



pBR328-Hinf 1

1.5% LPA, 6% LPA-coated
TBE pH 8.3
20 kV, 13 μA
L_{eff} = 5.6 cm, i.d. =75 μm, BF = 3
Inj: –10 kV, 3 s
260 nm (Ref 350 nm)
Capillary: 25°C
Carousel: 10°C

Fig. 3. Separation of PCR products using capillary gel electrophoresis.

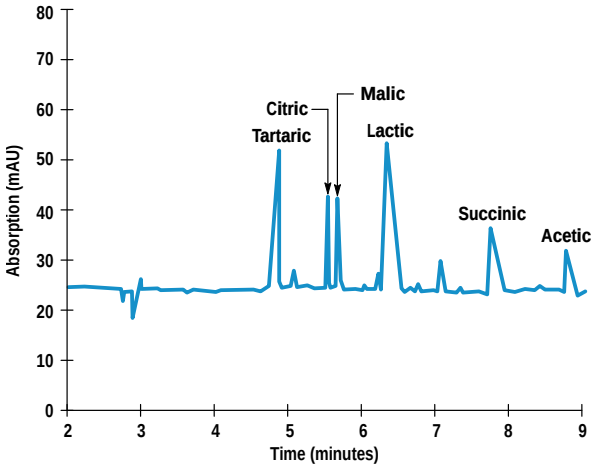
Table II
CE Applications and Benefits

Application	Other Analysis Methods	CE Benefits
Chiral Analysis	HPLC, GC, TLC, SFC	Speed Easy Method Development Cost of Analysis
Peptide Mapping	HPLC	Speed Orthogonal Mechanism
DNA Analysis	Slab Gel Electrophoresis, HPLC	Superior Resolution Speed Online Quantitation

All of these applications have in common that CE offers significant benefits over previously existing techniques (see Table II).

The future outlook for CE is positive although further development of capillaries suitable for protein analysis under native conditions and the development of other detection modules such as CE-MS will be important for long-term establishment of the technique.

Martin L. Verhoef
Product Manager
Waldbronn Analytical Division
David N. Heiger
Application Chemist
Analytical Marketing Center, Little Falls



Buffer: 5 mM phthalate, 0.25 mM CTAC, 0.07% β-CD, pH 3.5
Sample: Sake (diluted 1:5 with water)
Capillary: L_{eff} = 56 cm, L = 64.5 cm, i.d. = 75 μm
Injection: 200 mbar · s
Temperature: 15°C
Field Strength: 390 V/cm, Reversed Polarity

Fig. 4. Analysis of organic acids in sake employing indirect UV detection.

- Peak Purity Measurement. The fact that a spectrum is characteristic for a certain substance can be used to measure peak purity. If the spectra sampled along the peak are identical, the peak can be assumed to be pure. If the spectra change along the peak, a second substance might have coeluted.

Even though diode array detection is a desirable feature, it would be unacceptable if the sensitivity were not competitive with conventional UV/Vis detectors such as single-wavelength detectors or variable wavelength detectors, which work with filter wheels, bandpass filters, or rotating monochromator gratings, thus providing sequential spectra. Peak widths in CE are inherently smaller than in LC. The time needed by a scanning variable wavelength detector for scanning through the full spectral range cannot be neglected. This makes these detectors less preferable for spectral identification in capillary electrophoresis. However, the sensitivity of monochromator-based detectors is viewed as the state of the art in CE.

To obtain the same level of sensitivity with the diode array based spectrometer of the HP CE system, special care was given to the optical design of the CE detector. As described in more detail in the article on page 20, the objective was to optimize the light throughput and therefore the light incident onto the photodiodes, which determines the lowest noise level achievable with an optical detector. To maximize dA/dc, where dA is the incremental absorption change and dc is the incremental concentration change of the fluid residing in the optical path, all light emitted by the lamp is focused onto the

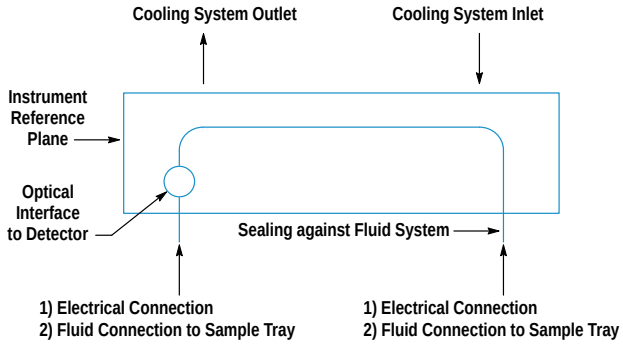


Fig. 3. Capillary interfaces to the instrument infrastructure.

capillary such that almost all rays pass through the inner diameter of the capillary and only a minimum of light (stray light) passes through the capillary wall without any interaction with the liquid inside the capillary. The benefits of this design optimization are:

- High light throughput. The result is an excellent noise level in the order of 2×10^{-5} to 5×10^{-5} AU (absorption units). This, combined with the low stray light level, provides superior sensitivity.
- Low stray light level. Besides excellent sensitivity, this results in a wide linear dynamic range, meaning that the detector responds linearly to an increasing sample concentration. The relevance of this feature can best be seen in a plot such as Fig. 4, which shows a synthetic polylysine preparation analyzed with CE.[1] The wide linear dynamic range makes it possible to quantify the smaller peaks (byproducts) relative to the main peak (active substance) very accurately.

The most remarkable feature of the detection system is the extended lightpath capillary, internally called the "bubble cell." By expanding the inner diameter of the capillary in the detection region by a "bubble factor" of BF (typically BF = 3) the sensitivity is linearly increased by a factor of BF according to the Lambert-Beer law,[2] assuming that the noise level of the detector is unchanged. This bubble cell makes the diode array detector one of the most sensitive optical absorption detectors. The larger inner diameter at the point of detection makes it possible to achieve a very low stray light level since more light rays can pass through the center of the capillary, thus increasing the linear dynamic range to 6000 (see page 23 for definition).

The HP CE diode array detector is also compatible with standard off-the-shelf capillaries. Different detector interfaces are available for the different inner diameters of capillaries with and without bubbles.

Liquid Handling

The purpose of the liquid handling functional module is to provide the necessary liquids to both ends of the capillary to facilitate a CE analysis. This includes electrolytes as well as samples, cleaning solvents, and waste vessels, as shown in Fig. 5. The module automates liquid handling so that the whole analysis can be done automatically and repetitively.

The two major design elements of the liquid handling module are the single-tray autosampler, which conveys vials to both ends of the capillary, and the replenishment system. The autosampler tray is designed so that it can position vials under both ends of the capillary independently and randomly. This enables the system to do analysis steps not possible with many other instruments.

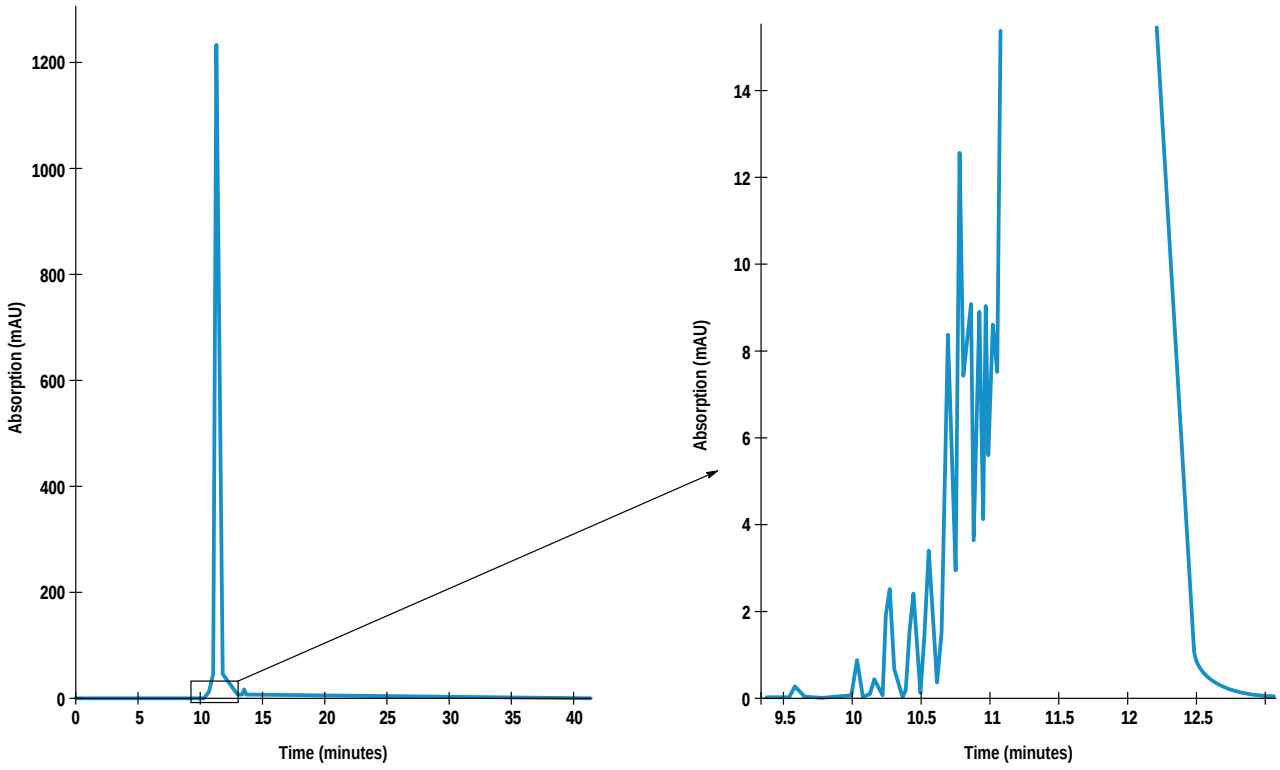


Fig. 4. Electropherogram of a synthetic polylysine preparation with 250 lysine residues.[1]

Random selection of the autosampler positions (there are 48 accessible positions) allows the user to select between sample, buffer, waste, and fraction collection vials. This gives high flexibility to the customer since the number of tray positions can be split according to the requirements of the application run on the instrument. When the same analytical method is being run in a repetitive mode, the buffer replenishment system can be employed and most positions can be filled with samples to be analyzed, thus providing maximum sample throughput.

Fractions of separated sample constituents are collected in a vial filled with a minimum volume of buffer. This is of particular importance for any further offline identification, for example with a mass spectrometer[3] or for further analysis with any other separation technique such as peptide/protein sequencing by Edman degradation.

Samples can be injected on both ends of the capillary. Usually an injection is done on the end farthest from the point of detection. However, for fast screening experiments it may be desirable to make a run on the short part of the capillary to save time. Reinjection from already collected sample fractions is also possible.

Buffer vials can be exchanged during analysis, independently on both ends of the capillary. For many applications it is mandatory to have this capability to complete a separation. For example, in isoelectric focusing mode the buffer change in pH can be used to transport the separated sample bands through the detector.

The replenishment system is primarily for the purpose of refreshing the buffers in the buffer vials by simply exchanging the contents of the vials with fresh buffer. This is required because the buffer undergoes chemical degradation by electrolysis which leads first to irreproducibility and in later stages to a shift of the pH resulting in a totally changed separation pattern. For unattended automated operation with high sample throughput, buffer replenishment is a mandatory feature, guaranteeing the highest reproducibility. Furthermore, the replenishment station adds other functions, including filling an empty vial to a selectable height or volume, emptying a vial to a selectable height, and sample or buffer dilution by adding buffer to a vial. This additional functionality also adds some basic sample preparation capabilities to the system. By ensuring the same liquid level in all vials, hydrodynamic flow effects are avoided, since no head pressure difference can occur.

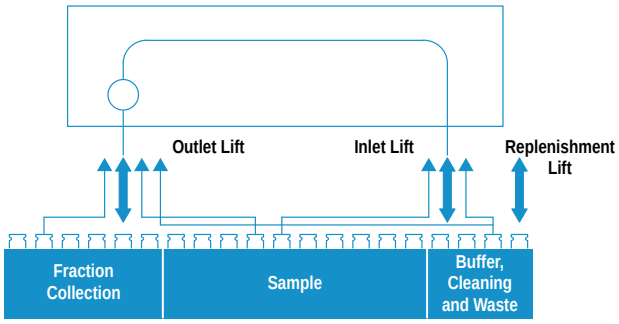


Fig. 5. The HP CE instrument liquid handling system includes an autosampler with random vial access from three lifts.

Pressure Injection

Because of the small volumes associated with CE, injection mechanisms employed in LC cannot be used in CE. For open capillaries (not filled with sieving matrices like gels) the most common injection mechanism is to apply a pressure differential across the capillaries. The volume injected follows the Hagen Poiseuille law:

$$\text{Volume} = \frac{\Delta p d^4 t}{128 \eta L}$$

where Δp is the pressure differential, d is the capillary diameter, η is the viscosity of the buffer, L is the total capillary length, and t is the time during which pressure is applied. For a given capillary the volume injected is basically the integral of the pressure-versus-time curve. The more precisely the pressure/time product is controlled, the more reproducible is the injection volume.

In the HP CE instrument the pressure/time product is controlled online and corrected immediately if any deviations are recognized from the programmed curve. Leaks in the system, which would result in a smaller injection volume, are automatically compensated by this procedure. The pressure is applied as a triangular function. On the downslope of this triangular pressure function, a vent valve is opened when the pressure is approximately equal to the ambient pressure. This has the advantage that the switching deviations of the vent valve have no significant impact on injection volume reproducibility. The pressure control algorithm significantly improves the reproducibility of the pressure injection system, as explained in detail in the article on page 50.

Pressure injection has some inherent advantages over injection by vacuum. The outlet of the capillary is not accessed during injection; this is important if other detection means like mass spectrometers are coupled to the CE instrument. Degassing and therefore air bubble formation in the capillary cannot occur, and fraction collection is much easier. Nevertheless, the design of the pressure system allows the user to apply a vacuum to facilitate injections from the detector end (shorter separation length), which may be important for fast screening experiments.

Electrokinetic Injection

Electrokinetic or electromigration injection is achieved by pulling the sample into the capillary by applying an electric field for a certain time. The quantity injected can be calculated.[4] This injection method is advantageous when viscous media or gels are employed in the capillary and when hydrodynamic injection does not work adequately. Basically, the volume injected is proportional to the charge introduced into the capillary. The HP CE instrument provides for the application of a voltage or a current for a certain time.

HP CE Technology Transfer

Looking at the new product generation process, often there is a long chain of activities involved, starting from basic research and proceeding to applied research, technology development, new product development, and finally, manufacturing. Later, while the product is being marketed, an improved and enhanced series of products will be generated in an R&D process sometimes referred to as current product engineering. Throughout this sequence of steps, knowledge, experience and competence are being built up, hopefully in a continuous, smooth, unidirectional way. However, since many people and different organizations are involved, losses, frictions, and interfacing issues associated with the transfer of know-how will occur and must be resolved or at least kept to a minimal negative impact.

Unlike the typical situation of 20 years ago, when the same group of people carried out the entire new product generation process all the way from research to manufacturing through many sequential steps, the scenarios of today's new product research and development projects have changed considerably. Given the complexity and breadth of most of HP's products, not only the analytical products, there is no way of making all of the key components entirely within HP. Instead, many things have to be acquired from external parties in various stages of development, be it fundamental research results, patented technology, methods, processes, or other know-how. In many situations even complete products or system components come from outside sources.

In our case of adding CE to the LPA (liquid phase analysis) product line, certainly we did look at the alternatives of acquisitions or external R&D collaborations, but before the final decision was made to take advantage of the accomplishments of an HP Laboratories research project and transfer technology and know-how from there to the Waldbronn Analytical Division and plunge into new product development, we had been thinking about mechanisms of technology transfer in general, and tried to reflect our conclusions in the organizational structure of our R&D function.

Generally, R&D divides its forces into activities of current product support and enhancement, development of next-generation products, and investigations, which include new products, new technology, key components, and fundamental research. The size of each of these segments depends on the business situation, which may change quickly. There is always a temptation to sacrifice the long-term investigations for short-term, market-driven problems or opportunities.

To stabilize the long-term, high-risk, but strategically important projects against the pressure of the tactical projects, we decided to divide the R&D function into two units: a smaller unit focusing on development, acquisition, and transfer of new techniques and generic components and a larger unit focusing on current and next-generation product development. The structure of the technology unit reflects the major technical and functional areas of the product line and therefore this unit has a few resident engineers and scientists who are specialists and experts in those categories. A larger number of members are set up in transient project teams and remain in the technology unit only during acquisition or investigation phases. We then transfer the project together with the transient team into the product development environment. This model should ensure minimal loss of know-how in the transfer from the investigation phase to the lab engineering phase and still keep a stable base of technical expertise and competence beyond one particular project cycle. In addition, it helps synchronize projects of different time scales and keeps the rules of the project life cycle flexible enough to adapt to new product design as well as to current product engineering.

The HP CE project was the first technology transfer project to be completed under this organizational structure and following these rules. The technology transfer from HP Laboratories to the Waldbronn Analytical Division yielded a very successful product. We have transferred technology from HP Laboratories before, but this transfer in particular went smoothly and pleasantly, I tend to believe, because of the new organizational model, but as much because of the enthusiasm, the dedication, and the support of the engineers and managers involved on both ends, including our marketing, manufacturing, and business people.

Alfred Maute
Engineering Manager
Technology Center
Waldbronn Analytical Division

The injection time table allows combinations of different modes of injection as well as flushing using randomly selectable vials. This makes it possible to do various sample manipulation steps that are essential for many applications. These include adding an external standard by injecting a plug of a standard compound, and adding a buffer with different conductivity to do sample stacking (enrichment) using the isotachophoresis effect.

Data Acquisition and Data Analysis

The software running on the PC next to the instrument has two major components: an instrument control module and a data collection and data analysis module.

Instrument Control. The functionality and versatility of the HP CE instrument made control by a PC the method of choice for handling the multiplicity of parameters and settings. The objective for the software design was to make the process of CE graphically visible and give the user continuous feedback on what the instrument currently is doing. As a design concept for the instrument control software an icon-based system was chosen. All instrument modules and their functions are symbolized. The user is shown a graphical representation of the instrument. Clicking on an icon displays a pop-up menu that offers the user the settings for that function. The entries are checked for plausibility and proper setting ranges. The HP ChemStation displays the status of certain events by changing colors, by showing the action taken, or by highlighting a setting the user has edited. The user can program the following parameters as functions of time:

- Replenishment
- Prerun and postrun conditioning
- Injection
- Voltage, current, and power
- Capillary temperature
- Pressure
- Inlet and outlet vial change
- Alarm limit (minimum limit for electrical current)
- Polarity of voltage or current
- Fraction collection.

For the detector, the following parameters are programmable as functions of time:

- Detector signals (defined below)
- Threshold and peak width
- Spectrum (permanent, all in peak, at up slope, peak maximum, and down slope, peak maximum, and baseline).

A feature of this software package is its simulation capabilities. It checks to see if the setup of a method to analyze a sample is in a logical sequence and if there are conflicting analysis steps such as, for example, a particular vial is needed at the inlet and outlet ends of the capillary simultaneously. When a simulation is started the graphical user interface reacts as it does during an actual run, allowing the user to follow and check online how the instrument will react in reality. Especially for complex methods, this feature helps prevent time-consuming correction steps and spillage or waste of valuable samples through wrong parameter or time table settings.

Different methods can be linked together to do a sequence of analyses. Several samples can be processed with the same method, or the same sample can be analyzed with different methods (different separation parameter settings), or a combination of both. Sequencing, together with the autosampler and the replenishment station, provides the instrument's automation capabilities.

Data Collection. The system control software collects raw data during an analysis from the following channels:

- Up to five detector signals (each signal includes wavelength, bandwidth, reference wavelength, and reference bandwidth)
- Spectrum (permanent, all in peak, at up slope, peak maximum, and down slope, peak maximum, and baseline)
- Voltage
- Current
- Power
- Capillary cassette temperature
- Pressure during run.

The first two items in this list are normally used to do the standard data analysis. For the detector signals, peak width and rise time are selectable according to the particular application. The last five data channels in the list serve the purposes of interpretation and documentation. In case of any unexpected results, this data can be used to localize and diagnose a particular problem with the instrument, the application, or the sample.

Data Analysis. When a method or sequence has been completed and the data is successfully stored on the PC, data evaluation can be performed by applying different algorithms. The information most relevant for the user consists of migration time, peak area, and peak height. When calibrated, the peak area represents a certain amount of sample. To determine the peak area, the detector signal is integrated by a built-in integration algorithm. Reproducibility evaluation of repetitive analysis with the same sample is often used to increase the reliability of the results. Besides integration, there are also built-in calibration and report generation functions. Calibration determines the unknown amount of a constituent of a known sample by calibrating with a predetermined concentration. Report generation provides different report style options including automatic spectral library searches.

One of the most powerful features of the system is its spectral capabilities. Spectra are measured using the diode array detector and stored as selected. The tools associated with spectra in the data analysis software are generally used to identify unknown compounds. The tools are:

- Spectral library. Allows a library search of selected spectra from a separation regardless of whether they are generated during an LC or a CE separation.
- Peak purity. Compares the ratio of two signals or spectra along the peak. If there is any deviation along the peak there is a probability of an overlay of two compounds.
- Isoabsorbance plot. This is for qualitative evaluation of spectral data. It can be used to generate a new electropherogram by extracting signals from spectral data, to find the optimum wavelength for detection of each component, and to average spectra to reduce noise.

Since identification by absorption spectra is a well-established tool in LC, many of the compounds analyzed with CE can be identified by available spectral libraries created with LC. The library shipped with the HP CE ChemStation software is identical to the library shipped with our LC ChemStation. This makes using the data analysis software more convenient for customers who already have experience with our LC ChemStation.

Summary

The HP CE instrument is the technical and technological implementation of the objectives stated for the project as well as for the product. Most of the requirements were implemented and can be reflected in a corresponding customer need. Significant improvements in detection sensitivity were made by the special detector design and even more by the extended lightpath capillary. The totally new liquid handling concept delivers significant automation features and versatility for future applications. The newly designed injection system provides further system reproducibility while the new forced-air-cooled cassette is a major step forward in terms of usability and flexibility. Another major advance for this type of instrumentation is the graphical user interface built into the HP ChemStation for instrument control. Finally, the electronic components and firmware developed for this product (discussed in the article on page 36) make this a functional product.

Acknowledgments

In the first place I want to thank Doug McManigill and his team from the analytical/medical laboratory of HP Laboratories for their outstanding contributions and commitment to capillary electrophoresis. To a great extent it is his achievement that HP has a product on the market. Special thanks also to Gary Gordon and his team, also from the analytical/medical laboratory, for their support and advice in the area of optical detection and for providing the crucial concept of the bubble cell. Thanks to Sid Liebes, Rich Tella, and Henrique Martins for their prompt help in the design and implementation of the bubble cell. My highest appreciation to the CE project team in Waldbronn; its commitment and dedication made it possible to finish the project on time. This includes a few individuals who were temporarily working on the project in Waldbronn, namely Henning Foukhardt, Rolf Dörrmann, Henry van Nieuwkerk, Greg Wilson, Monika Dittmann, and Sally Swedberg. Credit also to Alfred Maute

(Continued on page 19)

Industrial Design of the HP CE Instrument

It is not always recognized, and consequently not acted upon, that the contents and significance of industrial design for the success of products and the image of the company have changed in the past few years. Products have not only a technical, but also an aesthetic function. Together with some other representative features, such as advertising, company buildings, letterheads, packaging, exhibition booths, and the like, product design determines a major part of a company's image.† Today, products must not only be reliable, efficient and user-friendly, but also look like it. Visible quality today is the main attractor in many cases. In particular, the leveling out and standardization of technical achievements, even on the highest levels, lead to outward appearance being the decisive factor in the purchasing decision process more and more frequently. Last but not least, products, which also represent the company concept, have to be in line with the identity of the company. Thus, what is important is not avant-garde industrial design, but the successful combination of innovative and traditional elements.

Industrial design, therefore, is a part of product quality. Whether we thereby achieve the improvement of quality of use or acceptance among a specific target group, or capitalize on the identity of the company, proper industrial design is a sales promoting product feature. This applies not only to consumer products but also to commercial products. Very often, the first look at something determines whether we continue dealing with it or leave it alone. The quality of product design should make product quality the focus of customer interest right at first sight.

Internal Architecture

The first step towards good industrial design consists in laying out the components inside the instrument. As early as the investigation phase of the HP CE instrument project, the engineers were prompted to think about the dimensions and forms of the components they were responsible for. Rough details were then used to create all components from cardboard in duplicate. In a joint creative session, the entire project team used the models to arrange the components in multiple ways, finally deciding on one alternative. Understanding for each other's problems was gained within the team at a very early project stage: airflow and thermal problems, safety concerns, cooling, ease of use for service staff and users (which the industrial designer is also responsible for), and the like. This process not only led to the project being faster and more focused, but also had a teambuilding character.

† Image means the company as perceived by its customers. Identity means the company as it really is. Ideally, image = identity.

The immediate effects of the internal layout on the user interfaces become obvious in the vials, bottles, and cassette, which can be exchanged easily and intuitively. This ease of use is supported by the design of the instrument exterior, which visually reduces complexity.

As the project proceeded, a model of the outside cover was built (see Fig. 1). The visibility of the product in the form of the model further enhanced identification with the project, even beyond project team borders.

Appearance

Industrial design is not entirely up to the industrial designer. Carefully balancing innovative design with company-specific design within the sense of the corporate identity on the one hand, and on the other hand, emphasizing product-specific features to the best advantage, benefit both the product and the company.

The outward appearance of the HP CE instrument is designed to achieve a number of objectives:

- Emphasizing system character with a view to the HP PC, since the HP instrument is controlled by a PC. If the outward appearance of these two components is harmonized, it not only suggests that they come from the same company, but much more important, that they easily communicate with one another. In our case, this is achieved by equal use of volumes and forms as well as by HP identity elements such as coloring (bottom: dark grey, top: light), radii, HP nameplate, and power switch (see Fig. 1).
- Emphasizing system character with a view to HP analytical systems. The HP CE instrument is often to be found side by side with other HP analytical instruments in a lab or as part of an analytical system. If the outward features of our analytical instruments are in harmony with one another, this suggests to the customer that a complete solution to a problem is available from one source. In the case of the HP CE instrument, this is achieved by using, in addition to the design elements already mentioned, HP Analytical Products Group typical design or user elements such as, for example, the baseplate image, the semicircular pushbuttons set off by coloring, equal textures, equally colored windows, doors always opening in the same direction, equal status LEDs, equal fonts, and the like. The constant repetition of these visual and functional characteristics of the user interfaces makes a major contribution to ease of use by recognition. The particular challenge is always to find new solutions for new requirements that are technologically feasible but also continue the HP design tradition, so the customer finds it easy to get along and immediately accepts the solution as an evolutionary step.

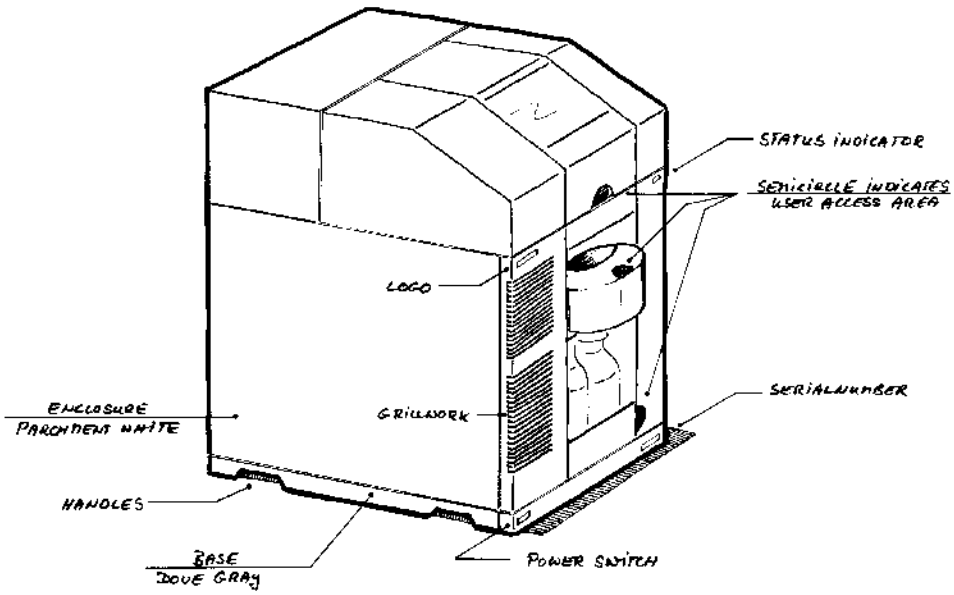


Fig. 1. Industrial design features of the HP CE instrument.

- The individual character of the HP CE instrument is emphasized by the basic instrument volume as well as by the visibility and form of some specific areas. The window parts show the areas relevant to the customer (cassette, sample tray, and replenishment bottles). Like these parts, the tray balcony, for example, is designed in such a way that it facilitates function (access to vials) on the one hand, while at the same time characterizing the look of the instrument clearly and unmistakably (and also optically reducing instrument depth). Emphasizing the vertical lines of the instrument optically supports the narrowness of the instrument.
- What is also particularly important is the visual expression of quality. Especially at a time when some instruments show only slight technological differences, customers should be convinced of the quality of our instruments at first sight. By using the right materials and manufacturing processes, we have attempted to distinguish our products qualitatively from those of our competitors, so that workmanship and finish provide our instrument with a highly professional appearance.

At the International iF Design Competition 1994, the HP CE instrument was awarded a prize for its good design and was seen by more than one million visitors in a special exhibition at the Hannover Fair and Cebit

Acknowledgments

The ideas of industrial designers often pose a challenge for those having to translate them into real parts. I would therefore like to thank all the managers for having strategically supported industrial design and, in particular, the mechanical engineers: Martin Bäuerle, Werner Schneider, and Hans-Peter Zimmermann, who despite occasional heavy time pressure made essential contributions to the quality of the industrial design.

Raoul Dinter
Industrial Designer
Waldbronn Analytical Division

for his technical advice and guidance during the project, to Lynn Yarke for her permanent support, particular during the early and uncertain phases of the project, to Manfred Seitz and his production engineering team, and to Wolfgang Wilde and his quality engineering team. Finally, it is a pleasure to acknowledge the contributions that Martin Verhoef, product manager for CE, and his team made to facilitate the commercial success of our product.

References

1. R. Grimm, *Comparison of Capillary Electrophoresis and HPLC for Checking the Purity of a Synthetic Polylysine Preparation,* Hewlett-Packard Company publication no. 12-5962-7230E.

2. A. Wiese et al, "A New Generation of LC Absorbance Detectors," *Hewlett-Packard Journal,* Vol. 41, no.4, April 1990, pp. 35-43.

3. M. Herold and Shiaw-Lin Wu, "Automated Peptide Fraction Collection in CE," *LC/GC,* Vol. 12, no. 7, July 1994, pp. 531-533.

4. D.N. Heiger, *High-Performance Capillary Electrophoresis—An Introduction,* Hewlett-Packard Company publication no. 12-5091-6199E.

A High-Sensitivity Diode Array Detector for On-Column Detection in Capillary Electrophoresis

The small peak volumes in CE demand special optical design to maximize sensitivity. High light throughput, good stray light suppression, and precise alignment are necessary. The diode array detector design focused on good matching of the illumination system and the spectrometer, precise alignment of the capillary and optical elements, and mechanical and thermal stability.

by Patrick Kaltenbach

Ultraviolet (UV) and ultraviolet/visible (UV/Vis) absorbance detection currently is the most commonly used detection technique in capillary electrophoresis (CE). The universal nature of this detection principle, its widespread use in high-performance liquid chromatography (HPLC), and its relatively simple adaption for on-column detection are some of the reasons for its popularity.

Using a diode array detector, absorption at a large number of wavelengths can be measured simultaneously, giving spectral information. Thus the area of application can be further enlarged. Since the observed spectrum is characteristic for any analyte, this information can be used for identification of unknown compounds. Multiwavelength detection, peak identification, and peak purity determination are capabilities that make the diode array detector a powerful tool and an ideal detector for CE.

The small peak volumes in CE demand special optical design to maximize sensitivity. High light throughput, good stray light suppression, and precise alignment of the small capillary are necessary to get the best possible performance.

Optical System

Fig. 1 shows a schematic overview of the optical system of the HP CE diode array detector. A new type of prealigned deuterium lamp is used to increase the overall light throughput and thus improve the sensitivity of the system. This type of lamp has a much smaller emission spot than lamps previously used for HPLC, making it superior for illumination of small sample areas. The polychromatic light of the lamp is focused onto the detection window of the capillary by an achromatic lens system. These detection windows are usually made by removing a small section of

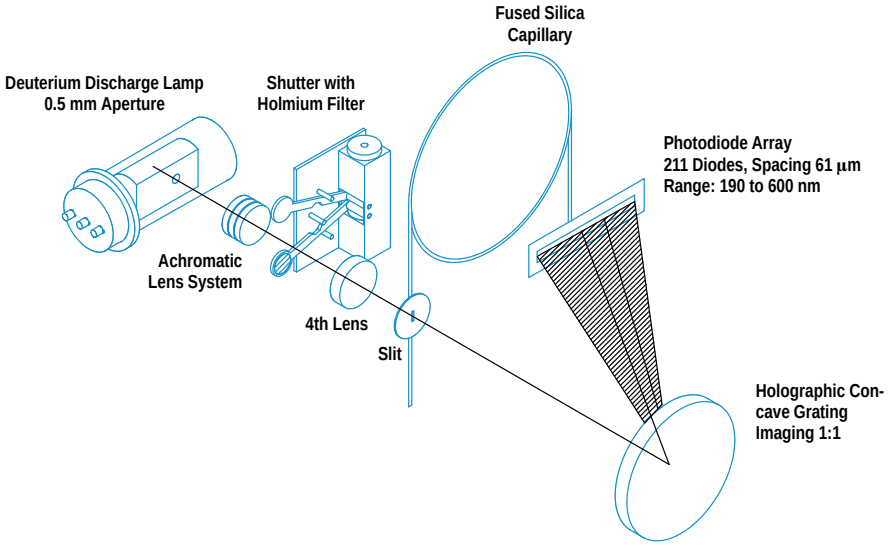


Fig. 1. Schematic overview of the CE diode array detector optics.

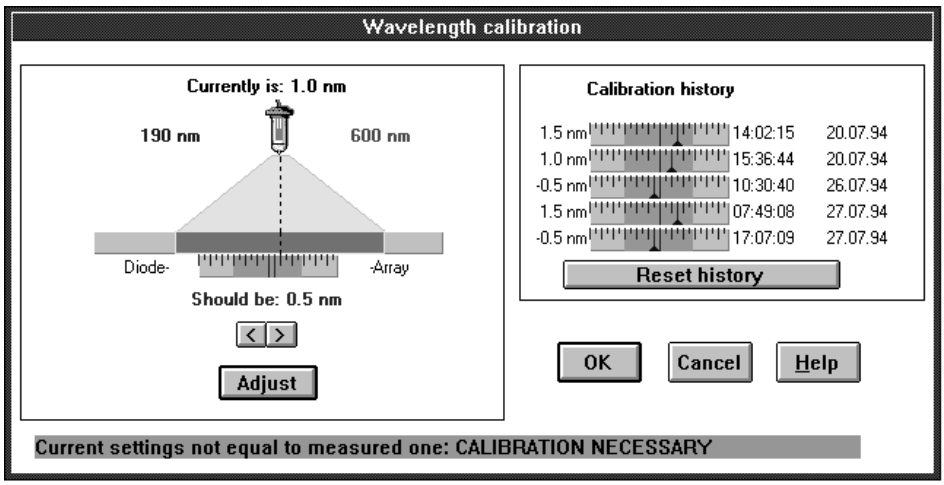


Fig. 2. Wavelength calibration window.

the polyimide coating of the fused silica capillaries. A slit behind the capillary blocks stray light and defines the entrance slit of the spectrometer. The shutter is used to cut off radiation for dark current compensation or to drive a holmium oxide filter into the optical path for wavelength calibration purposes. The spectrometer consists of a holographic concave grating which disperses the light onto a photodiode array. 205 diodes are used for a wavelength range from 190 to 600 nm, which gives a resolution of 2 nm per diode. Grating, array, and signal electronics are very much the same as for the HP G1306A diode array detector and the HP 79854A multiwavelength detector for HPLC.[1]

To optimize the overall performance of the system, the design focused on good matching of the illumination system and the spectrometer, precise alignment of the capillary and optical elements, and mechanical and thermal stability. The optics bench is precision machined and exchangeable parts such as the lamp or the alignment interface (slit) are self-aligning. To maintain optical precision, the system recognizes when the user accesses these exchangeable parts. A wavelength calibration is performed every time the instrument senses a possible change and if necessary the user is asked to confirm new calibration settings (Fig. 2).

Since the spectrometer f-number of 4 represents the basic limit on the light throughput of the system, the lamp and lenses are designed to match this limitation to optimize the overall light throughput.

Ray tracing software was used to calculate aberrations, to determine theoretical optimums for positioning of the optical elements, and to maximize light throughput. Fig. 3 shows a ray tracing plot of the optical system, and Fig. 4 shows in detail a ray trace through a capillary with 150-μm inside diameter—for example, a 50-μm capillary with an extended lightpath with BF = 3 (BF = bubble factor, the increase in the diameter of the capillary at the detection window of the detector, as explained in the article on page 62).

Theory

On-column detection approaches often use setups similar to that shown in Fig. 5. In a small area near the outlet the capillary is illuminated with light from a UV/Vis light source, and a slit of width o is placed behind the capillary to block stray light through the capillary wall. The light transmitted through the slit is detected. Using the Lambert-Beer law, the detector signal for an ideal cell is determined by:

$$A(\lambda) = e(\lambda)CD = \log\frac{I_0(\lambda)}{I(\lambda)},$$

where $A(\lambda)$ = absorbance in absorbance units (AU) as a function of wavelength, $e(\lambda)$ = extinction or molar absorption

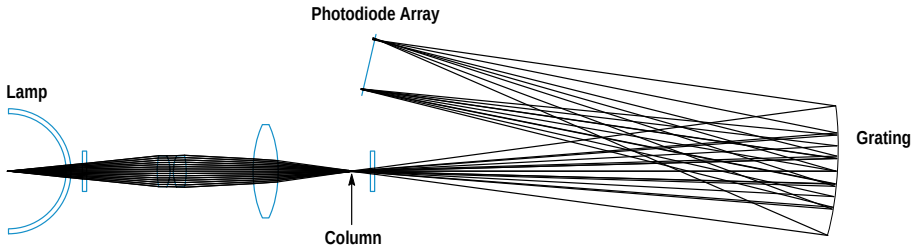


Fig. 3. Raytrace of wavelengths 190 and 600 nm through the CE diode array detector optics.

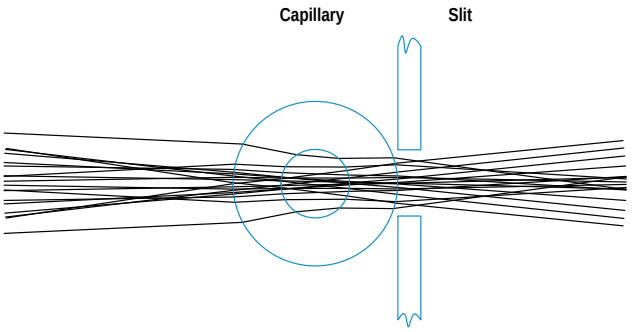


Fig. 4. Raytrace of 190, 254, and 600 nm through a water-filled capillary.

coefficient as a function of wavelength, C = molar solute concentration, D = path length, I_0 = incident photon flux, and I = transmitted photon flux.

If stray light passing through the column wall can reach the detector, the equation has to be rewritten:

$$A = \log\frac{I_0 + I_S}{I + I_S},$$

where I_S = stray light through the capillary wall.

A typical detector response is shown in Fig. 6. The lower limit of detection is determined by the baseline noise of the detector and its sensitivity, which can be described by the slope dA/dC of the detector response. The upper limit of detection is determined by the nonlinearity of the detector response at higher concentration levels. Unwanted stray light causes a deviation of the response from the theoretical linear slope according to the Lambert-Beer law. The steepness of the slope, in contrast to typical HPLC detection cells, is dependent on the effective path length (related to the inside diameter of the capillary) and the stray light through the capillary wall. As a result, there is a linear range in which the detector can be operated to get reliable results.

Sensitivity, Noise, and Linear Dynamic Range

This section discusses the parameters that influence the performance of the diode array detector and the trade-offs that have to be made to optimize the system.

A big disadvantage of the on-column detection approach is the possibility of light passing through the transparent capillary wall without interacting with the sample. Figs. 4 and 5 show clearly that not all rays propagate through the center of the circular cross section of the capillary. This decreases the linearity of the detector and reduces the effective path length, which determines the sensitivity of the detector. To maximize the response (effective path length) and linearity of the detector, the width o of the slit must be as small as possible.

Since the slit defines the entrance aperture of the spectrometer, the slit width o has an impact on the optical bandwidth. A 60-μm slit width corresponds to a 2-nm optical bandwidth of the CE diode array detector.

At the lower limit of detection, baseline noise becomes important. As long as the baseline noise of the signal is limited to Schottky noise (noise associated with the photocurrent generated by the photodiodes), the baseline noise is dependent on the total light throughput reaching the detector and is inversely proportional to the square root of the light throughput (photocurrent):

$$I_{S\,p\text{-}p} = \frac{6\sqrt{2qI_{ph}BW}}{2.3I_{ph}},$$

where $I_{S\,p\text{-}p}$ = peak-to-peak value of the Schottky noise current, q = charge of the electron, I_{ph} = photometric signal (current), and BW = electrical bandwidth.

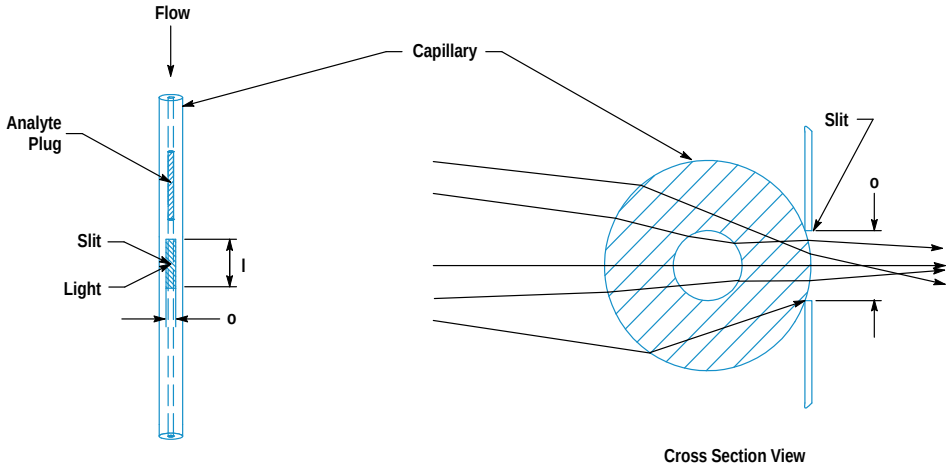


Fig. 5. Slit and capillary in an on-column detection approach.

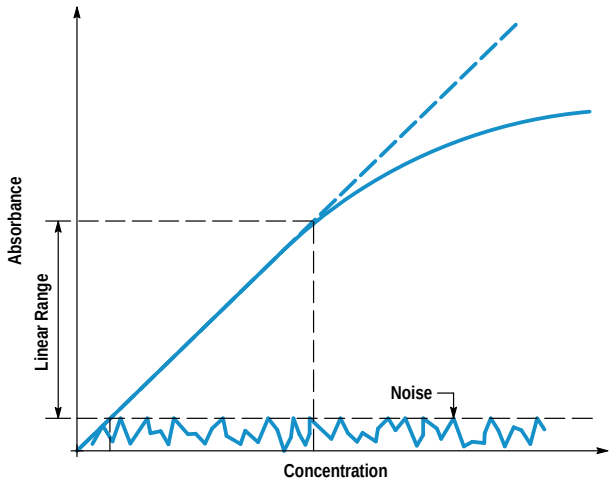


Fig. 6. Diode array detector response characteristic.

Expanding the slit, therefore, will decrease the baseline noise. Thus the optimum slit width o is a trade-off between light throughput (noise), linearity, sensitivity, and spectral resolution. An important attribute for detector performance is the linear dynamic range, calculated as:

$$\text{Linear Dynamic Range} = \frac{\text{Upper Limit of Detection (AU)}}{3 \times \text{Noise (AU)}},$$

where the upper limit of detection is given by the point where the detector response deviates 1% from the expected linear response (see Fig. 6).

Different capillary inside diameters are used in CE (typically 25 μm to 75 μm). For optimum performance the slit size must be well-matched to the inside diameter. Using the right slits, a high linear dynamic range can be achieved for standard capillaries and even further improved using extended lightpath capillaries (Fig. 7).

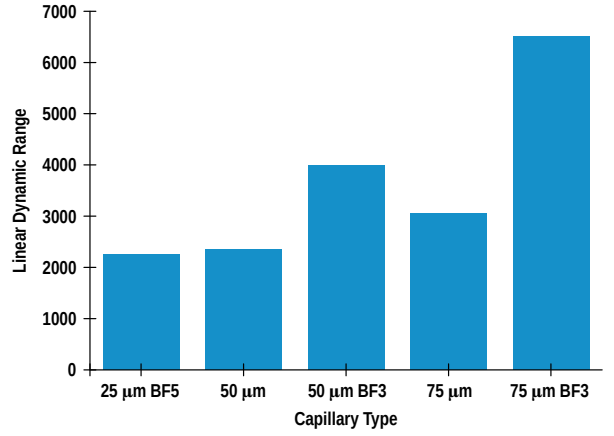


Fig. 7. Typical values for the linear dynamic range using different capillary inside diameters. BF = bubble factor, the increase in the diameter of the capillary at the detection window of the detector.

The slit height l along the separation axis has an influence on the resolution (peak width) and peak height of a detected peak. It also has an impact on the light throughput and influences the baseline noise as already mentioned. To minimize the impact of the detector on the efficiency of the separation the slit height should be as small as possible. Fig. 8 shows the theoretical distortion of a Gaussian peak detected through a rectangular slit. For simplification, the mathematical model neglects the separation (mobility) of the analyte (peak) in the detection window. The deformation is given by:

$$\text{Signal Deformation} = \frac{\text{Observed Signal}}{\text{True Signal}},$$

where the observed signal is the detector output using a slit with height l, and the true signal is the detector output using an infinitely small slit.

Longer slits result in a loss of resolution (broader peaks) and decreased peak height of the recorded signal. However, as long as the slit is small enough compared to the peak width, slightly longer slits do not significantly degrade resolution and they improve the signal-to-noise ratio by allowing more light to reach the photodiodes. As a rule of thumb, the slit height should be smaller than the parameter σ of the detected peak, where the Gaussian peak shape can be described by the formula:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{\sigma^2}}$$

$$\text{Peak Width} = \sqrt{5.54}\,\sigma \text{ (full width at half maximum).}$$

HP offers a variety of capillary lengths and inside diameters to cover the needs of a wide range of applications. Three standard inside diameters are available at the moment: 25, 50, and 75 μm. The 50-μm and 75-μm capillaries are offered with and without extended lightpaths (bubbles), whereas the 25-μm capillaries are only offered with a five times expanded inside diameter at the detection point (BF = 5).

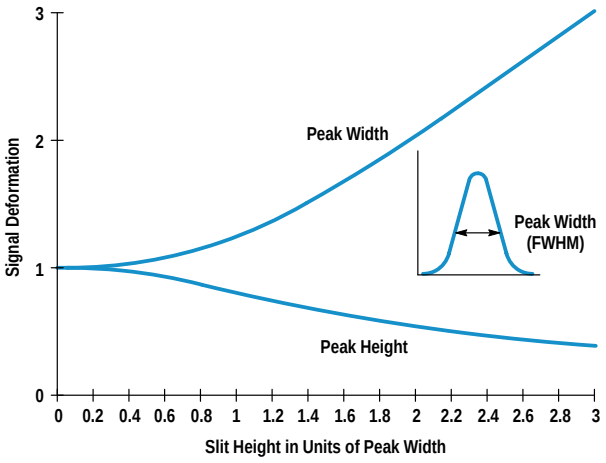


Fig. 8. Signal deformation by increasing the slit height.

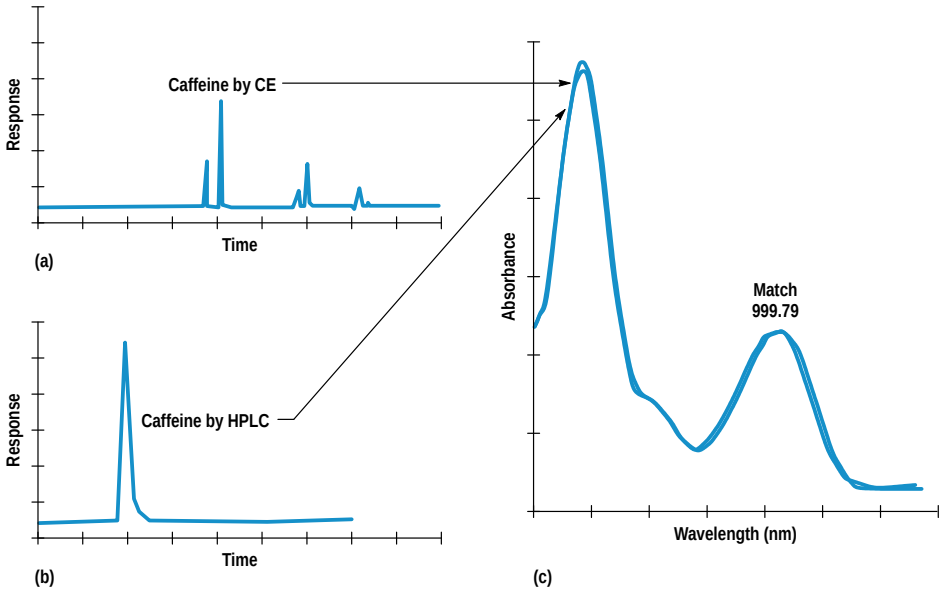


Fig. 9. Comparison of caffeine spectra acquired by (a) CE and (b) HPLC. (c) Overlay of caffeine spectra from CE and HPLC runs. CE conditions: buffer = 11.7 mM borate, 8.3 mM phosphate, 50 mM SDS, pH 8.9; Sample = acetaminophen, caffeine, naproxen, guaifenesin, phenacetin, noscapine; Capillary i.d.= 50 μm, L_{eff}=56 cm; injection pressure/time product = 50 mbar·s; E =370 V/cm; T = 37°C; detection wavelength/bandwidth = 200/16 nm. LC conditions: Hypersil ODS column, 100×2.1 mm, 5-μm particle size; T = 40°C; mobile phase = water/acetonitrile (85/15) at 0.4 ml/min; 10 ng caffeine on column; detection wavelength/bandwidth = 254/4 nm; slit width = 2 nm.

Table I shows the dimensions of the slits that are offered for each capillary type. The minimum useful signal bandwidth indicates the related optical bandwidth.

Table I
Capillary Inside Diameters and Slit Dimensions

Capillary Inside Diameter	Slit Dimensions (o×l)	Minimum Useful Signal Bandwidth
25 μm, BF = 5	120×80 μm	4 nm
50 μm	40×620 μm	2 nm
50 μm, BF = 3	145×145 μm	5 nm
75 μm	55×620 μm	2 nm
75 μm, BF = 3	200×150 μm	7 nm

BF = Bubble factor, the increase in the diameter of the capillary at the detection window of the detector.

Spectral Quality

Capillary electrophoresis is becoming more and more important as an orthogonal technique to HPLC. The use of spectral libraries for peak identification, using spectra taken in HPLC, can be very helpful.

However, the optical design of the CE diode array detector must ensure that no distortion occurs because of different effective path lengths for different wavelengths. The ray trace in Fig. 3 indicates that rays of different wavelengths may travel different paths through the cell resulting in different effective path lengths. Special care has to be taken to minimize distortion; otherwise, spectra will differ from those taken with an HPLC system.

As an example, Fig. 9 compares the spectrum of caffeine acquired with a 50-μm standard capillary on the HP CE diode array detector with that obtained from an HPLC system equipped with a 2-nm slit. The match is as good as typical matches obtained on the same instrument with repetitive analysis.[2]

Acknowledgments

I would like to acknowledge the valuable contributions of Henning Fouckhardt, who set the directions for this project during the investigation phase. Special thanks to Bernhard Dehmer for his great support and advice during the entire project, and to Rolf Dörrmann for the casting design.

References

1. A. Wiese, et al, "A New Generation of LC Absorbance Detectors," *Hewlett-Packard Journal*, Vol. 41, no. 2, April 1990.

2. D. Heiger, et al, "Diode Array Detection in Capillary Electrophoresis," *Electrophoresis*, November 1994.

Capillary Handling in the HP Capillary Electrophoresis Instrument

Capillaries are encased in cassettes for easy replacement and connections are made automatically when a cassette is installed. Air cooling of the capillary eliminates leak problems and lowers costs. Vials containing samples and electrolyte are automatically lifted from a tray to either end of the capillary.

by Hans-Peter Zimmermann

As the name capillary electrophoresis (CE) indicates, the separation capillary is a fundamental element of a CE instrument. Depending on the application, different types of capillaries are used. Capillaries can differ in length (from 25 cm to 100 cm), internal diameter (25 μm to 150 μm), or coating of the inner walls. Thus the user must have the ability to exchange the capillary easily and quickly.

In the HP CE instrument, this ability is provided by a capillary cartridge system. The capillary is encased in a cassette. By keeping several cassettes on hand, each with a different type of capillary, the user can exchange the capillary in the instrument and be ready for a different application within a few seconds. The cartridge system makes it possible to change the capillary in the cassette in a few minutes should the capillary be blocked or break, or should its surface be deactivated. Fig. 1 shows the cassette setup.

Using the Cassette

The cassette is sealed with a quick lock mechanism. The user only needs a rollerball pen to open the cassette (Fig. 1a). The capillary is threaded through the detector interface (Fig. 1b), which later, upon placement of the cassette, adjusts the capillary to the detector. The capillary can now be inserted in the cassette and reeled up on the mandrel. A clamp mechanism fixes the inlet capillary end in the cassette. The two feet of the cassette facilitate aligning both capillary ends to the same length while at the same time protecting them. When needed, the cassette is inserted into the holding fixture provided in the instrument and tipped back. Fig. 2 shows the cassette location in the instrument.

When the user inserts the cassette into the instrument, the following interfaces are established automatically:

- To the high-voltage power supply for generating the electric field within the capillary
- To the sample and electrolyte handling system (liquid handling system) to provide the capillary with sample and electrolyte for automatic operation
- To the pressure system to be able to push liquid into the capillary for injection or flushing purposes
- To the detector for measuring the absorption of the substances separated in the capillary
- To the capillary cooling system for thermostatically controlling the separation capillary temperature.

Connection to the High-Voltage Power Supply, Pressure System, and Liquid Handling System. Upon insertion of the cassette, the capillary ends are automatically threaded through the

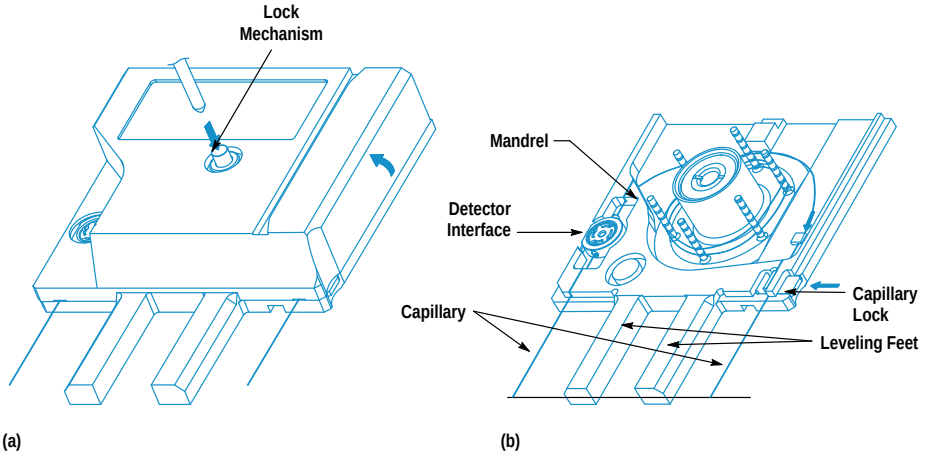


Fig. 1. Capillary cassette. (a) Opening the cassette. (b) Inserting the capillary.

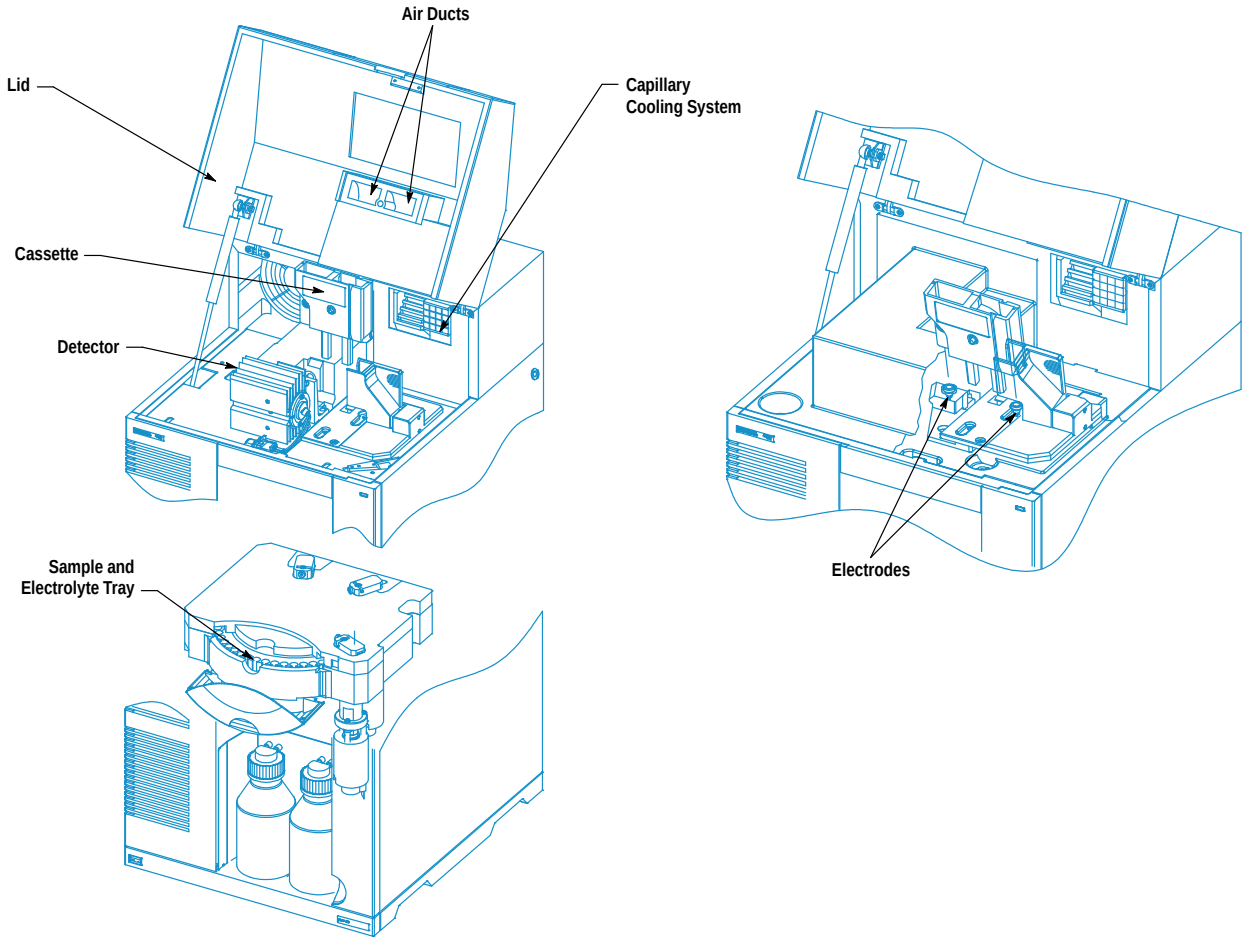


Fig. 2. Inserting the capillary cassette.

two electrodes located in the instrument. The electrodes establish the links to the high-voltage power supply, the pressure system and the liquid handling system.

Fig. 3 shows the setup of an electrode. The funnel threads the capillary through the silicone rubber seal, the housing, and then through the platinum tube electrode. This electrode is connected to the high-voltage power supply and makes the electrical connection to the two electrolyte vials during analysis to apply the electric field for separation to the capillary. To avoid corrosion of the electrodes caused by the electric field and corrosive buffers, the electrode tubes are made of platinum, which has extreme chemical stability. For sample analysis, vials containing the sample or the electrolyte buffer are lifted up to the two electrodes so that the capillary end and the enclosing platinum tube dip into the liquid.

To flush the capillary or to inject sample, it is necessary to seal the separation capillary against the vial, so that pressure can be applied inside the vial. If the lid of the instrument is closed, a spring-loaded pin inside the lid presses the cassette down on the electrodes. This compresses the silicone rubber plug inside the electrode housing so that the capillary is sealed against the electrode. During flushing, a vial is sealed to the cone at the bottom of the electrode

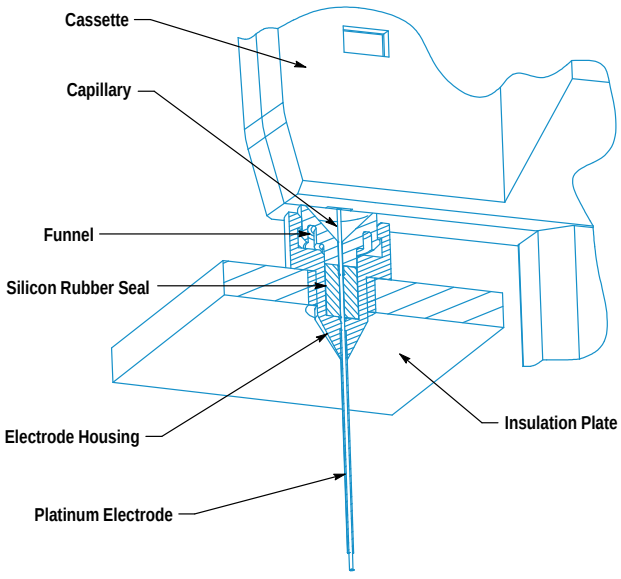


Fig. 3. Electrode details.

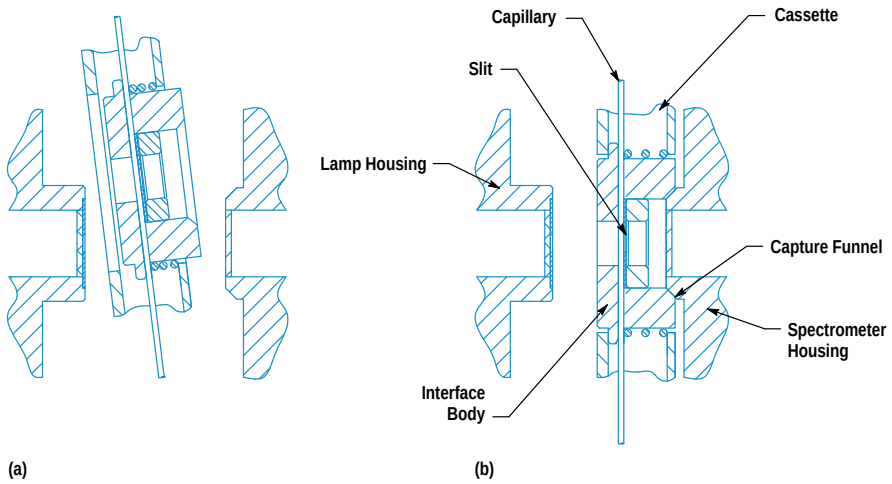


Fig. 4. Detector interface. (a) Sliding in the cassette. (b) Final position.

housing so that the pressure inside the vials presses liquid into the capillary.

Connection to the Diode Array Detector. When the capillary is threaded through the detector interface, it is aligned radially by a V-shaped slot to the slit in the detector interface, which shades the capillary and leaves only the detection window open. The correct axial position of the capillary relative to the slit is achieved by a small stopper affixed to the capillary. The size of the slit depends on the internal diameter of the capillary, so different detector interfaces must be used for the different capillary types. The correct choice of interface for each type of capillary is indicated by a color code.

The detector interface floats in the cassette with a clearance of ±1 mm. When the cassette is tipped back, the detector interface is automatically adjusted to the detection slit by a capture funnel (see Fig. 4).

Connection to the Capillary Cooling System. For cooling and temperature control of the separation capillary, air is blown into the capillary cassette at high speed. The capillary cooling system is located in the upper back part of the instrument (see Fig. 2). When the lid is closed, air ducts in the lid automatically link the cooling system to the cassette. The function of the capillary cooling system is described in more detail in the next section.

Capillary Cooling System

Depending on the applied field strength and conductivity of the buffer located in the capillary, there is an energy throughput of up to 6W in the capillary during analysis. If the capillary were cooled only by natural convection, the liquid core in the capillary would rise to more than 80°C above room temperature in a 1-m capillary at maximum power. This might cause destruction of sensitive samples or even boiling of the buffer. Fig. 5 shows the setup and the different capillary and fluid temperatures T_c, T_1, T_2, T_3, and T_f.

The temperatures in the capillary can be calculated according to Newton's Law: For the temperature difference in the core, the following is true:[1]

$$\Delta T = T_c - T_3 = \frac{P}{2\pi L}\left(\frac{1}{2k_b} + \frac{1}{k_s}\ln\frac{r_2}{r_1} + \frac{1}{k_p}\ln\frac{r_3}{r_2}\right), \quad (1)$$

where r_1, r_2, r_3, and L are the dimensions of the capillary (see Fig. 5), k_b, k_s, and k_p are the thermal conductivities of the buffer, the quartz wall, and the polymer coating, and P is the power level inside the capillary.

Using typical substance values and measures, equation 1 predicts a temperature difference between the core and the outside of the capillary of about 2.0 degrees with a power of 4.5W and a 1-m capillary length. The temperature gradient within the capillary only plays a minor part. The main contributor to the temperature rise in the core is the heat transfer to the ambient air at the capillary outer surface. This temperature difference can be calculated according to the following equation:[1]

$$\Delta T = T_3 - T_f = \frac{P}{2\pi L r_3 h}, \quad (2)$$

where h is the heat transfer coefficient at the capillary surface.

Unfortunately, the temperature drop at the capillary surface is not as easily calculated as equation 2 suggests, since the heat transfer coefficient h has a nonlinear dependence on the flow velocity and the temperature of the surrounding coolant. Fig. 6 shows the temperature rise in the core (T_c–T_f) as a function of flow velocity for air and for a typical cooling liquid (fluorocarbon FC77 from 3M) with a power dissipation of 4.5 W/m in the capillary.

Fig. 6 shows that when cooling the capillary with forced air, above a flow velocity of about 10 m/s, improvements in cooling performance cannot be achieved. For liquid cooling, the temperature rise inside the capillary is about 3.6 times lower than for air cooling at equal velocity, according to equation 2. Flow velocities as high as with air, however, can hardly be

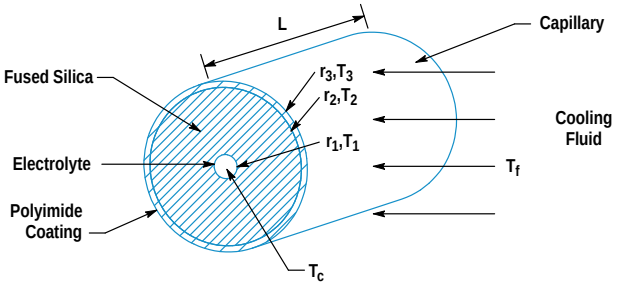


Fig. 5. Temperatures and radii of fused silica capillaries.

Rapid Prototyping for the HP CE Project

In the past few years, methods offering the possibility of transforming 3D CAD data directly into a ready-to-install part have experienced a steep upturn. These methods, which have been commercially available for about ten years now, are generally known under the catchword "rapid prototyping." Some of these methods and their combination with different replication techniques have reached the real aim of rapid prototyping, namely the reduction of time to market. Within the HP capillary electrophoresis (CE) project, rapid prototyping methods became indispensable, not only because of the ambitious project schedule, but also because of the fact that some parts, because of their complexity, could not have been produced in the traditional way without substantial compromises.

At the Waldbronn Analytical Division, these methods were used for the first time, involving significant risk because none of the people concerned could rely on personal experience. Another reason for using rapid prototyping methods certainly was the cost aspect. However, it proved necessary to balance the requirements of the individual development steps, the time available, and the complexity of the individual parts against one another. In accordance with part requirements, three methods were pursued in the HP CE project, two ways to get a resin-plastic part and a third method to get an alumina part:

- 3D data → stereolithography process → vacuum casting → plastic model
- 3D data → Solider process → vacuum casting → plastic model
- 3D data → STL file (Fig.1) → stereolithography process →STL model (Fig. 2) → vacuum casting with silicone → wax model (Fig. 3) →ceramic shell for traditional investment casting for alumina parts (Fig.4).

The real rapid prototyping step results in directly translating the 3D construction data into a solid part. The principle of all three methods mentioned above is similar: a 3D computer model is split into layers of a certain thickness by software.

Various methods are applied to copy the section in question to materials such as photopolymers or powdery thermoplastics by using different reproduction techniques. This procedure is repeated until the whole part, composed of hundreds of layers, is complete (see Fig. 5).

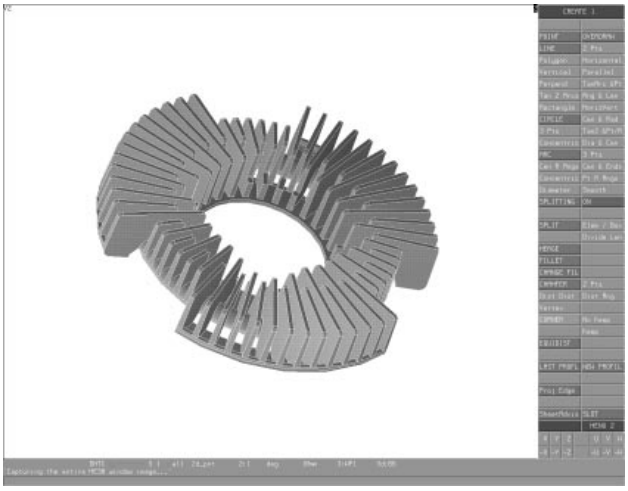


Fig. 1. STL file created with a CAD application.

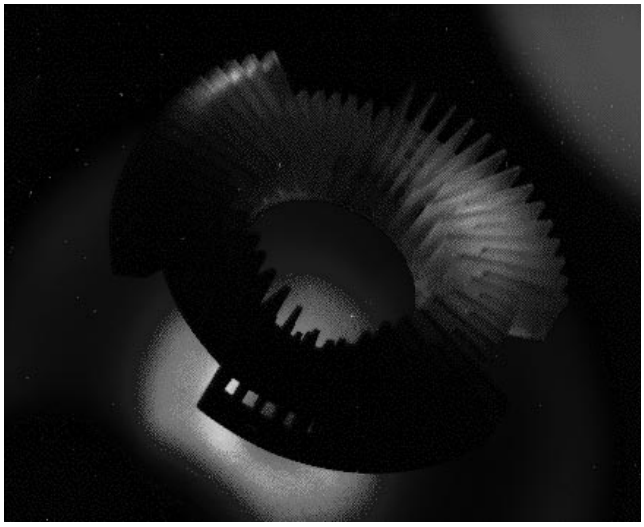


Fig. 2. STL model.

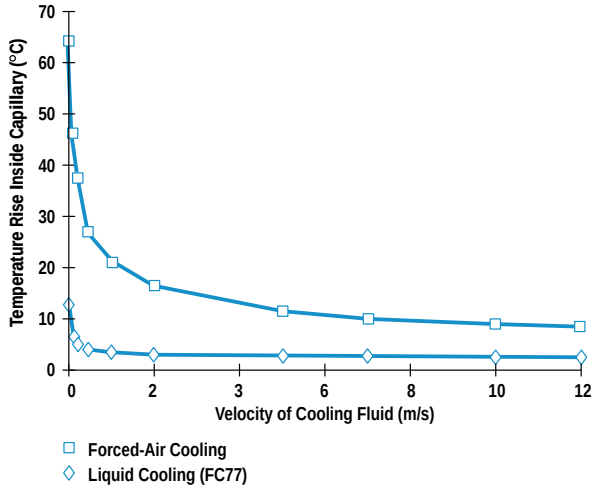


Fig. 6. Temperature rise inside capillary.

realized with liquids. At typical velocities of about 0.5 m/s with liquid cooling, the difference from high-speed forced-air cooling only is about 5°C (8.8° for air; 3.9° for liquid) with a 4.5 W/m energy throughput in the capillary. The advantage of air cooling versus liquid cooling is that capillary handling is significantly easier for the user. When exchanging the capillary, no liquid has to be drained and no costly sealing of the cassette is needed. With air as the coolant, low leakage has no disturbing influence. Also, no expensive cooling fluids are necessary, which lowers the cost of ownership.

Because of these advantages of air cooling, we decided that an excess temperature of five degrees in the capillary core could be accepted. In the HP CE instrument, forced-air capillary cooling is used with air velocities of about 10 m/s at the capillary.

Because the mobility of the substances in the capillary changes by about 2% per degree, another important requirement of the cooling system is the ability to adjust the temperature of the cooling air with a precision of ±0.1 degrees to suppress variations in room temperature. This is achieved by a Peltier heat exchanger element regulated by three temperature sensors. This element makes it possible to adjust

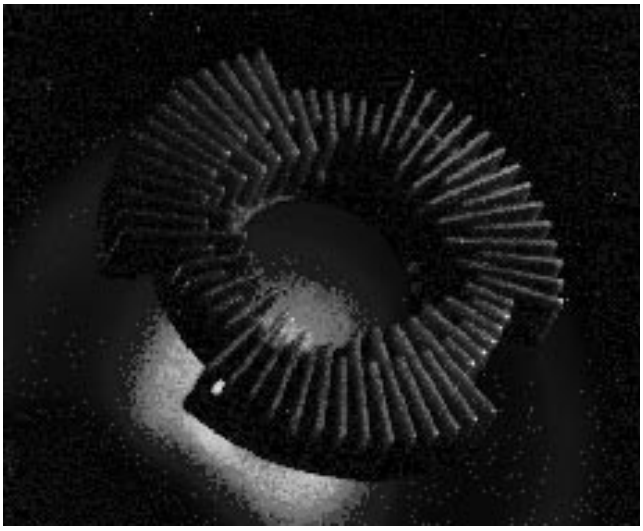


Fig. 3. Wax model.

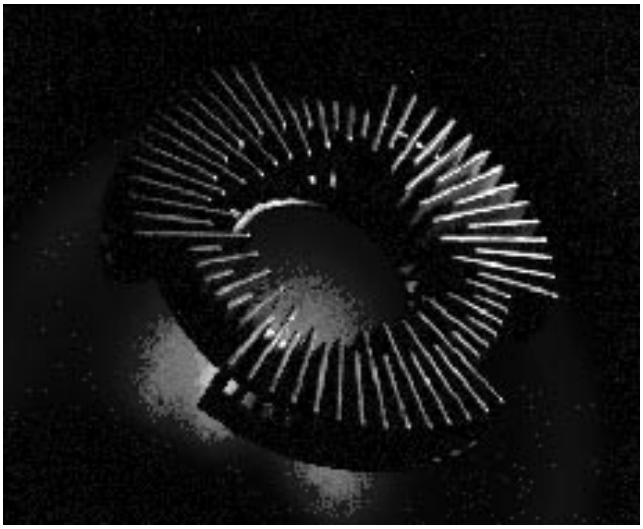


Fig. 4. Production part.

Depending on the method applied, the 3D data has to be prepared in different ways (e.g., support framework). Regarding the actual formation of the part, however, each method has its process-specific advantages and disadvantages.

In qualification of these rapid prototyping methods, however, it must be pointed out that the parts cannot be used without costly subsequent treatment and replication—preferably with silicone—if specific requirements on surface quality, accuracy, and stability are to be met.

Martin Bäuerle
R&D Engineer
Waldbronn Analytical Division

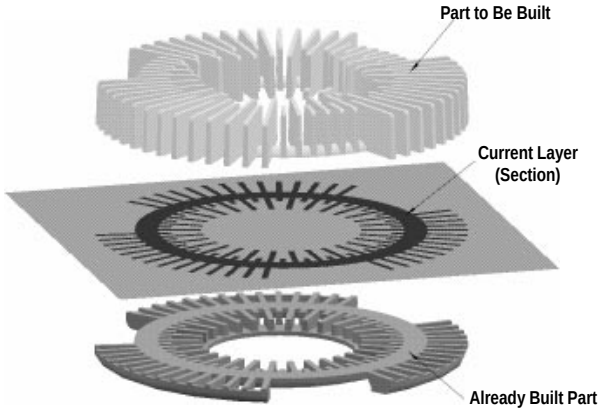


Fig. 5. The solid models are built up in layers.

the air temperature from ten degrees below ambient to 60°C. Fig. 7 shows the setup of the cooling system.

A radial blower draws air past the heat exchanger of the Peltier module and blows the cooled or warmed air towards the cassette through the air ducts in the lid. From there, the air is fed back to the heat exchanger. The Peltier element is able to deliver up to 50W of cooling power and up to 100W of heating power to the capillary cooling circuit. Because of the relatively low efficiency of a Peltier element, there is a power dissipation of up to 150W at the heat sink of the Peltier module. To remove this heat, the heat sink is forced-air cooled by an axial ventilator. Ventilator, Peltier, heat exchanger, and driver electronics are located in the cooling system, embedded in two isolating foam shells.

Sample and Electrolyte Handling

To achieve a high sample throughput in a CE instrument, the supply of samples and electrolytes to the two capillary ends has to be automated. In the HP CE instrument, this task is performed by the liquid handling system shown in Fig. 8.

Samples and electrolytes are in small vials of 2 ml or 0.3 ml volume in a circular tray ring. The tray ring can accept a maximum of 48 vials. A heat exchanger in the center of the tray ring can be connected to an external cooling bath for adjusting the vial temperature. The heat exchanger is an example of several parts of the HP CE instrument that were designed using rapid prototyping, as explained above. The tray ring rests on three rollers at its perimeter and is driven by a servo motor via a pulley.

Samples and electrolytes in the tray have to be positioned at both the capillary inlet and the capillary outlet. Because the tray ring must be free to rotate, the vials must be lifted from the tray to the capillary ends. This task is performed by the lift modules. Two lift modules lift the vials up to the capillary ends, and a third lift module moves vials to a needle

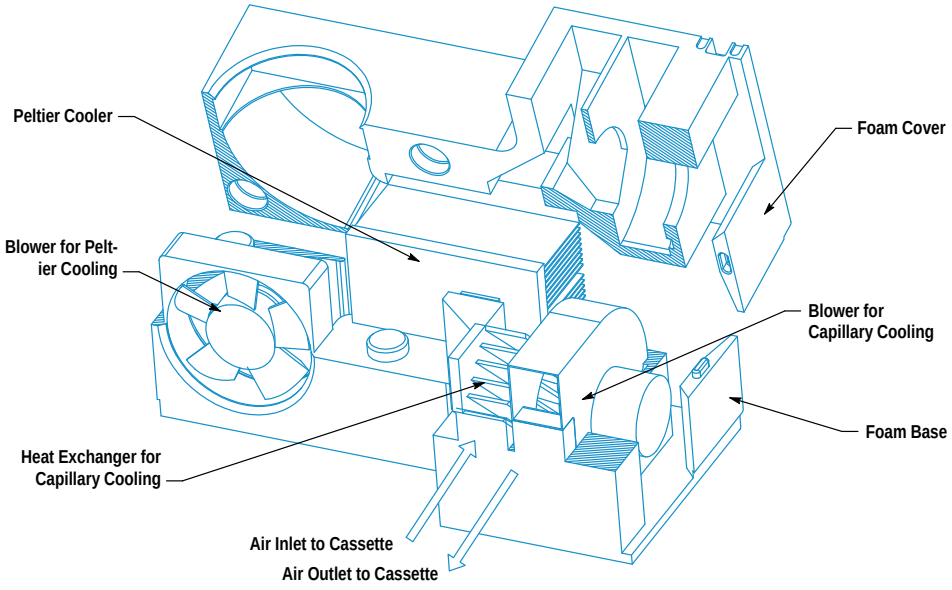
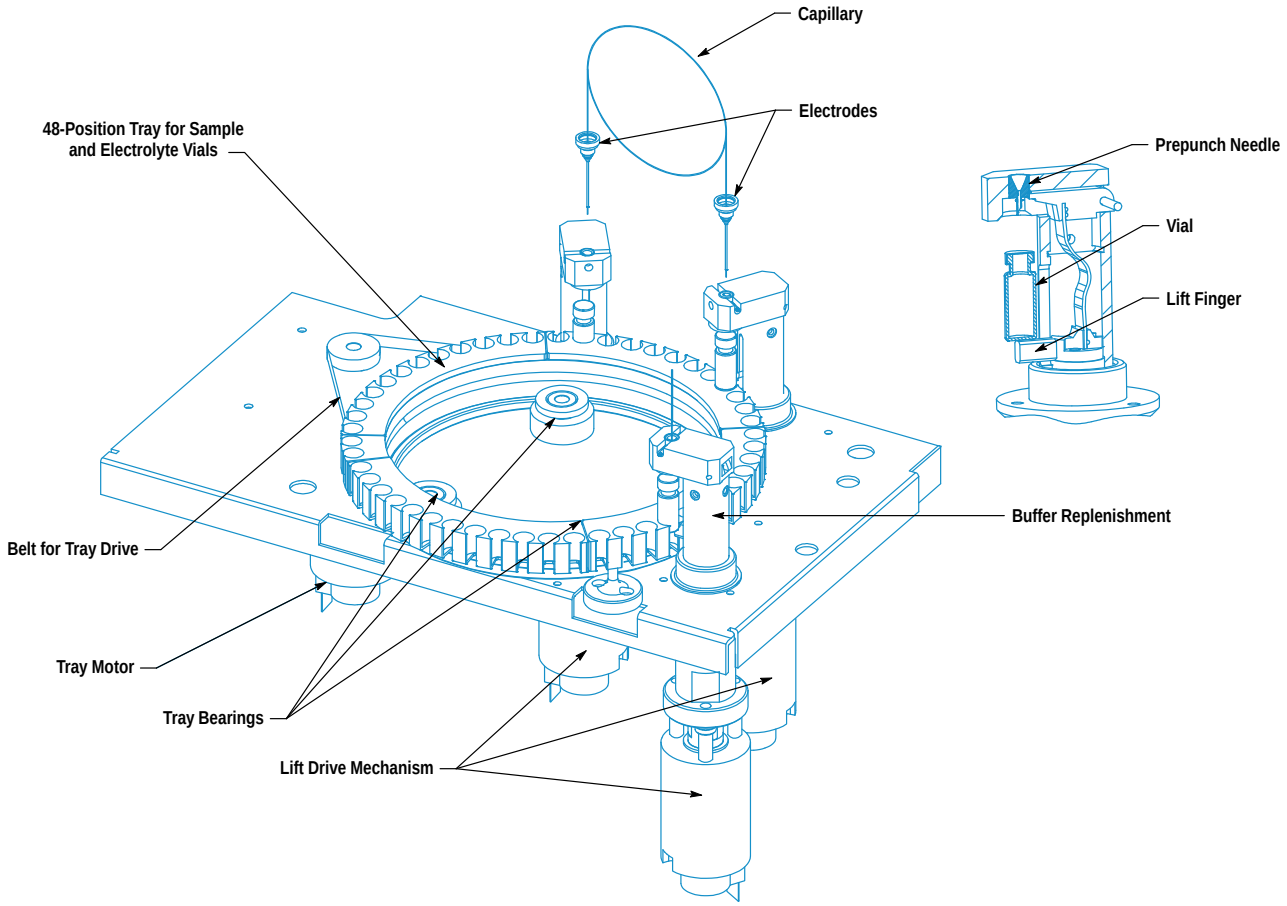


Fig. 7. Cooling system.

through which contaminated electrolyte can be sucked off and vials can be refilled with fresh electrolyte from a dispenser bottle (see article, page 32).

This setup offers several advantages. First, it provides random access to all of the vials in the autosampler. Each vial can be positioned either at the capillary inlet, at the capillary outlet, or at the replenishment module. This means that the user can decide how much of the tray capacity to use for sample vials and how much for electrolyte vials. Second, changing vials at the capillary outlet allows the different fractions of a sample to be collected separately. Third, this approach allows buffer replenishment and capillary preconditioning to proceed in parallel.

To avoid sample evaporation, the vials are sealed with caps whose thin membranes are punched upon injection. The sensitive ends of the separation capillary would break in this



ig. 8. Liquid handling system.

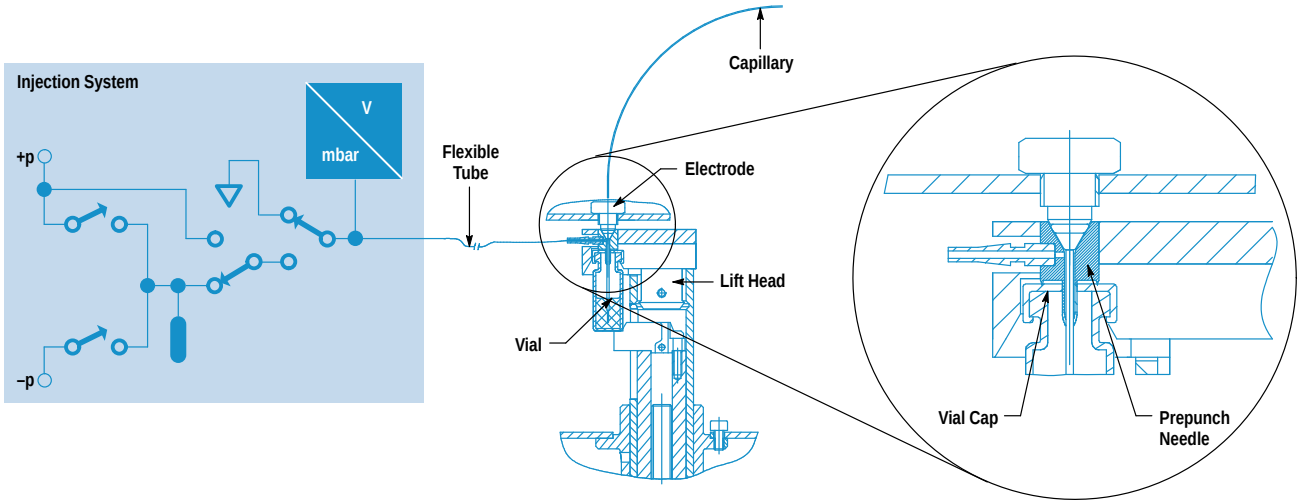


Fig. 9. Connection to the injection system.

process, so a punching needle is located in the lift head to prepunch the caps. In operation, the vial is held by a kind of pliers and is pressed onto the punching needle by a spring before being lifted to the capillary. A ring-shaped blade seals the prepunch needle to the vial cap. The lift presses the cone of the prepunch needle against the seal at the electrode housing (the seal can be seen in Fig. 3), thereby providing the seal between the vial and the capillary. Pressure for flushing or injection can then be applied to the vial through a bore in the prepunch needle. The final position of the vial is shown in Fig. 9.

Acknowledgments

I would like to acknowledge the contributions of John Christianson, who set the directions for the capillary and liquid handling techniques in the HP CE instrument. Also, special thanks to Rolf Dörrmann and Greg Wilson for their support during the project.

Reference

1. J.H. Knox and K.A. McCormack, "Temperature Effects in Capillary Electrophoresis, Part 1: Internal Capillary Temperature and Effects upon Performance," *Chromatographia*, Vol. 38, no. 3/4, February 1994, pp. 207-221.

Sample Injection in HP CE

For flushing or conditioning the capillary or injecting a sample, air pressures or different values and durations are applied. The injection system provides precise closed-loop control of the integral of the air pressure over time for either direction of fluid flow. The replenishment system automates the exchange of used electrolytes for fresh ones, using a special double-needle design.

by Werner Schneider

When analyzing samples with an electrophoretic instrument, there are times when the liquid contents of the separation capillary have to be moved, such as for flushing the capillary with different fluids for cleaning purposes, for conditioning it for another separation run, or for introducing sample in a quantifiable and reproducible manner. In the HP CE instrument, these tasks are performed by applying air pressures of different values and durations. How this is done is the subject of the first part of this article.

The second part of this article describes the replenishment system. During a separation run, the chemical properties of the working electrolytes change as a result of electrochemical reactions. This is typically not a problem for a single run, but for multiple runs the run-to-run stability or repeatability will be lost. The replenishment system provides the user with an automatable function for exchanging the used electrolytes with fresh ones.

Overview

Fig. 1 shows a schematic diagram of the functions mentioned above. The pressure source provides working pressures. The EMI board handles the output of the four pressure sensors (indicated by the arrows out of the other blocks) and drives the solenoid valves (the two, three, or four arrows shown going into the other blocks). For more details about the relationship of these functions to the other functional areas of the instrument, especially to the system control, see the article on page 36.

Pressure Source

Flushing, injection, and replenishment are driven by gas pressure. To avoid the need for external gas supplies, the pressure source is built into the instrument. It consists of a small dual-acting membrane pump, two valves (intake and outlet), an intake air filter, and two pressure sensors (for pressure and vacuum). Dual-acting means that the pump provides both overpressure (+p, around 930 hectopascals or millibars) and vacuum (–p, around 420 mbar below ambient). The outputs of the valves are connected to the two large bottles, which serve as pressure "capacitors" and as containers for the fresh electrolyte and the waste. Thus the containers are also parts of the replenishment system. The firmware controls the pressure values with two-point closed-loop control, with only one of the two valves active at a time.

Flushing and Injection

Driving a liquid through a tube is in general fairly straightforward: immerse the ends of the tube in liquid, apply a pressure at the inlet end, and you get a flow. According to the Hagen-Poiseuille law, the flow rate is proportional to the pressure and to the fourth power of the internal diameter of

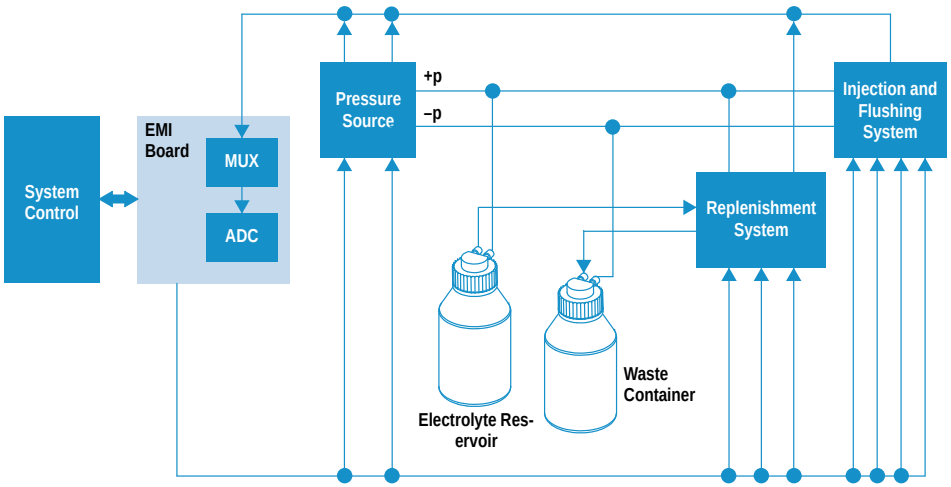


Fig. 1. Overview of the injection and replenishment systems.

the tube, and inversely proportional to the length of the tube and the viscosity of the liquid.

Flushing a capillary means applying a relatively high pressure to achieve acceptably high flow rates. As a rule of thumb, at least three to five times the liquid contents of the capillary have to be pumped through to achieve a thorough flush. A typical flush time is two minutes for a capillary with a 50-μm inside diameter and a length of 48.5 cm.

Injecting a sample means applying pressure and driving the liquid into the capillary. Here only small amounts of liquid are brought into the capillary, and the repeatability of the sample volume must be high. The run-to-run variation must be in the range of 3% or better.

Small amounts of liquid means 1% to 2% of the capillary volume (or, in more common CE terms, 1% to 2% of the capillary length). Since all the factors of the Hagen-Poiseuille law can be considered constant in the short term, the flow rate is directly proportional to the pressure difference across the capillary. This means that the integral of the pressure over time is an acceptably precise measure for the amount of liquid forced into the capillary. Since the user wants to select the injection amount as a parameter of the method, the instrument needs to generate the time integral of the pressure. For example, with a 50-μm inside diameter, 48.5-cm-long capillary, a pressure of 40 mbar for 4.6 s achieves a sample plug approximately 3 mm long.

Fig. 2 shows the schematic diagram of the injection and flushing system, together with a view of the inlet lift holding a sample vial. The lift is part of the liquid handling system described in the article on page 25. Valve 1 and valve 2 deliver air flows for injection, and valve 3 is for flushing. Valve 4 is used for controlling the venting port. Venting is active during all movements of the components to avoid the buildup of any spurious pressures. For example, some compression takes place when the O-ring between the lift head and the electrode holder has begun to seal but has not yet reached its final sealing position.

Flushing. Flushing is straightforward. After the vial is fetched and brought to the sealing position, both valve 3 and valve 4 are activated. Valve 3 directly connects the vial to the system pressure +p, pushing the liquid into the capillary. When the user-defined time has elapsed, both valves are switched off again, and the lift head is released from the sealing position to its default height for the analysis run.

Injection. As indicated above, injection requires much smaller pressures than flushing. For avoiding detrimental effects like splashing, the pressure is not applied instantaneously but in a smooth manner. The manifold that carries the valves also has an internal chamber. When switching valve 1 on, for example, air from +p flows through restriction 1 into this chamber. With the sample vial in the sealing position and valve 4 activated, the air flow causes a rise in the pressure in this closed system (phase 1 in Fig. 3), which is measured by the injection pressure sensor. This is analogous to an electrical resistor-capacitor arrangement, that is, the pressure rise follows an exponential curve. Since the maximum value of 50 mbar is small compared to that of the pressure source, the rise is almost linear. The same mechanism is applied to generate the downslope. Through valve 2 and restriction 2, the closed system is connected to −p, thus drawing the air out of the closed system (phase 2 in Fig. 3).

The timing of the valves is controlled by a successive approximation algorithm implemented in the instrument firmware. Actively controlling both the upslope and the

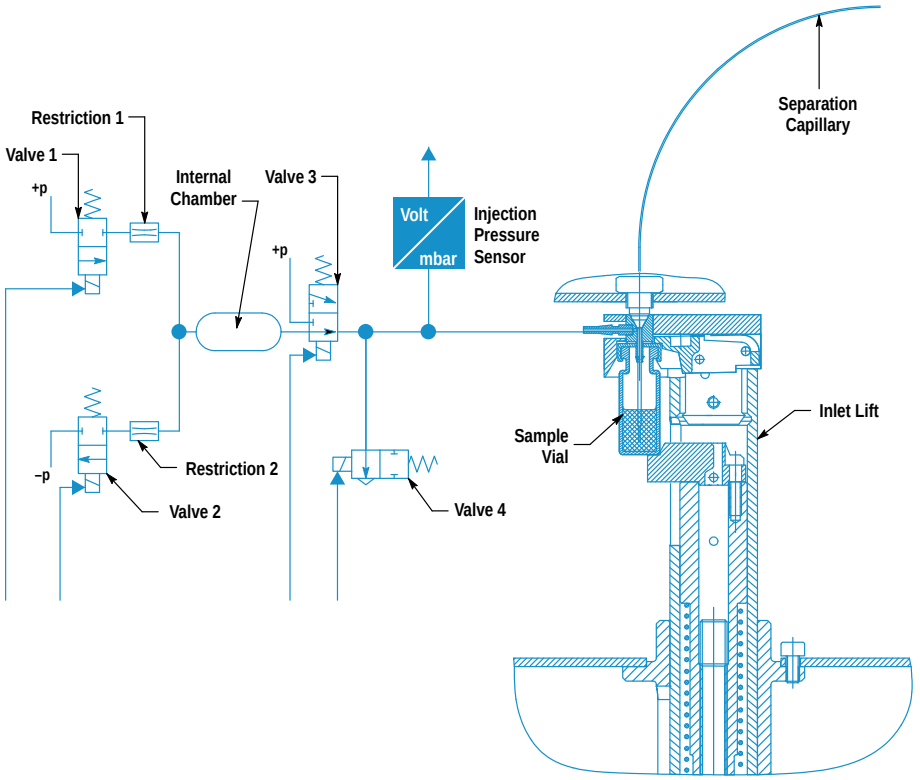


Fig. 2. Injection and flushing system.

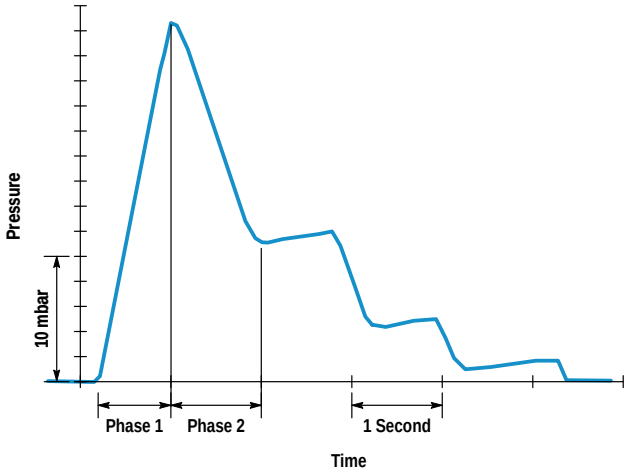


Fig. 3. Capillary pressure profile during sample injection.

downslope provides precise closed-loop control of the integral of the pressure over time. Its precision and reproducibility are determined by the timing resolution of the instrument's multitasking system and the resolution and accuracy of the pressure sensor. The system totally compensates for the variations and tolerances of the other components, including the switching characteristics of the valves, the tolerances of the restrictions, the temperature dependences of dimensions, and gas viscosities.

The total area under the pressure curve in Fig. 3 represents approximately 50 mbar·s. Values down to 20 mbar·s (this corresponds to a plug length of 0.3 mm) are achievable with less than 0.3% variation. Hence the variations of the injection scheme are significantly smaller than those caused by the so-called "zero injection effect" as described in the article on page 50.

Additional Features

The bipolar pressure source allows not only closed-loop control of the injection cycle, but also generation of a similar curve with reversed sign, that is, the flow direction can be reversed. This is advantageous for separations performed from the detection side of the capillary, such as fast screening separations.

By appropriate firmware control, this scheme also allows the application of small pressures in the range of −50 mbar to +50 mbar. Like the other method parameters, this function is time programmable and is available during the separation run.

For separation methods like isoelectric focusing, where no electroosmotic flow is present, external pressure is necessary for pushing the separated species past the detector.

Replenishment

Repeated runs for every sample are mandatory for valid analysis results. Therefore, it is necessary to keep all the conditions and parameters that influence the separation as stable as possible. In electrophoretic analyses, the electroosmotic flow is one of the major aspects, and is strongly influenced by the pH value of the run electrolyte. As a result of the voltage applied to perform the separation, electrochemical by-products are generated. One of the effects is the decomposition of water. This leads to the generation of protons at the positive electrode and hydroxide ions at the negative. More protons in a solution means that the solution gets more acidic, and more hydroxide ions do the opposite. These changes directly influence the electroosmotic flow. In general, the electrolyte solution is buffered to suppress this change, but after a number of separation runs, the buffer capacity is exhausted. In the HP CE instrument, one of the key features is automated, unattended operation. Instead of using valuable vial positions in the instrument's autosampler for a large number of run buffer vials, the replenishment system offers the ability to work with a minimum set of buffer vials, supplying fresh electrolyte into them. This provides the user with the maximum number of positions for sample vials.

The replenishment system is shown in Fig. 4. It consists of the electrolyte reservoir, the waste container for holding the used liquid, valve 2 for drawing the used liquid out of the vial into the waste container, valve 1 for dispensing fresh electrolyte into the vial, a measuring scheme for controlling the whole process, and a double needle for accessing the vial.

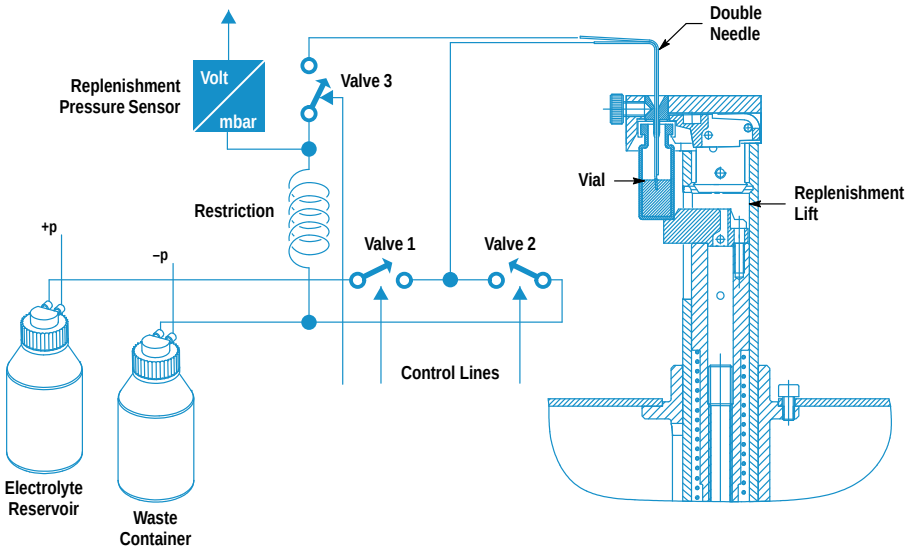


Fig. 4. Replenishment system.

The double needle is provided by the replenishment lift and serves for emptying and filling the vial and as the probe for the measuring function. It consists of two pieces of stainless steel capillary which are welded together with an axial offset of their orifices of five millimeters. The double needle is attached to the frame of the instrument. The lift moves the vial up and down and the double needle enters the vial through the prepunch needle in the lift head (see article, page 25).

The measuring scheme uses the pressure sensor, the restriction, the measuring line of the double needle, and valve 3. It detects when liquid comes in contact with the orifice of the measuring line, that is, when the shorter end of the double needle touches the liquid surface. When the measuring system is activated by switching valve 3 on, a small flow of air is drawn through the measuring line and the restriction. Like a pair of electrical resistors, these two parts of the flow path act like a voltage divider. The ratio of the restriction values is approximately 1:15. With one end of the restriction connected to −p and the open end of the measuring line at ambient pressure, the connection point is at approximately −25 mbar. This value is measured by the pressure sensor. As soon as liquid touches the orifice, the air flow is interrupted. Thus, the pressure at the connection point is drawn towards −p. The control algorithm implemented in the instrument firmware interprets this change of the pressure value as the signal for "liquid detected." The related parameters, especially the air flow and the internal volumes of the lines and the pressure sensor, are selected so that the delay between touching the liquid and issuing the signal is less than 100 milliseconds.

The vial is filled and emptied through valve 1 and valve 2, which are connected to the longer end of the double needle. Switching on valve 2 applies the vacuum. This is done before the double needle is immersed in the liquid. The lift then slowly raises the vial while the liquid is drawn into the waste container. The free end of the double needle is tilted at a slight angle to avoid blocking the orifice when the bottom of the vial is reached.

With valve 1 activated and valve 2 closed, liquid is dispensed into the vial, driven by the pressure in the electrolyte reservoir. The filling height is selectable through the positioning function of the lift. It is determined by the sensing scheme described above and is user-programmable as part of the method for every vial independently. For the highest repeatability of the results, the user might replenish the electrolyte for every run. To keep the consumption of electrolyte low, the user might select the minimum filling level (the electrodes must be immersed during the run). In separations that are chemically robust, the buffering capacity of highly filled vials might suffice for six to ten runs.

When preparing the separation capillary for the next run, in the preconditioning phase of the method, the capillary is flushed with buffer. These flushes consume liquid, so from time to time the buffer vial has to be filled up again. For this purpose, the system offers the refill function. Instead of emptying and filling, the vial is only refilled to the desired level.

It is possible to perform parallel replenishment by using two pairs of run buffer vials and replenishing the pair of vials for one separation run while another separation run is in progress. This means that after the initial setup, no additional time is required for replenishment, provided that the separation run time is not extremely short.

Acknowledgments

Among all the people who were involved with this work, I would like to explicitly mention Fritz Bek who designed the firmware, Franz Bertsch for the electronic design, Klaus Witt for his support and contributions, Hans-Peter Zimmermann for his valuable input to the double-needle concept, and Fred Strohmeier for his guidance and patience during the development of the instrument.

HP CE Separation Control Electronics and Firmware

The HP CE instrument consists of a PC and a base unit consisting of detection and separation subunits. Methods are developed on the PC and downloaded to the base unit for independent execution. The control electronics and firmware of the separation subunit takes care of tray and vial movement, capillary voltage, current, and power control, capillary temperature control, diagnostics, and related data capture.

by Fritz Bek, Franz Bertsch, and Klaus Witt

The HP capillary electrophoresis instrument is built of two hardware units. One is a personal computer and the other is the base unit, which contains all of the hardware needed to perform the separation and detection. The electronics and the instrument software (firmware) of the base unit are the subjects of this article.

The software running on the PC is the only interface to the user. There is no local keyboard on the base unit. A three-color LED is used as a status indicator on the front of the instrument.

The PC (an HP Vectra computer) is connected via the HP-IB (IEEE 488, IEC 625) to the base unit and performs all data handling and user interface tasks. The base unit is built as two subunits—the separation subunit and the detection subunit—and therefore has two HP-IB addresses.

There is no logical connection between the subunits. This allows the separation subunit to be compatible with future advances in detection technology and makes it possible to implement other detector types without changing the separation unit.

The detection subunit (see article, page 20) is very similar to the multiple wavelength HPLC detector of the HP 1050 LC system.[1] From the control point of view it is identical to the HPLC detector and electronically it is exactly the same. The cardcage, power supply, and printed circuit boards are unchanged. The local user interface is deleted.

Like the detection subunit, the separation subunit has a cardcage in the rear that holds the various printed circuit boards and the power supplies. Most of the printed circuit boards are connected to a motherboard on the backplane of this cardcage, except for the temperature driver board, the status indicator board, and the electromechanical interface board, which includes pressure sensors, connections to Hall sensors, and the power outlets for the various solenoid valves. The electromechanical interface board extends from the front of the subunit to the motherboard. The mechanical hardware for liquid handling is connected by cables to this board or the motherboard.

Controlling a Method

A *method* is the sum of all of the parameter settings and commands that control a complete analysis. A choice of activities can be programmed to:

- Refill liquid into working vials (replenishment)
- Prepare the separation capillary (preconditioning)
- Inject a certain amount of liquid from the sample vial (injection)
- Perform the separation and collect data (run time)
- Evaluate data and generate reports (data analysis)
- Clean up the system (postrun and postconditioning).

During electrophoresis, electrolytic byproducts are formed at the electrodes in the inlet and outlet vials, and the sample compounds move to either end of the capillary. Therefore, the running buffer will be spoiled after use. To prevent influencing later analysis results, replenishment is used to prepare several vials with fresh buffer. Any existing liquid in the vials can be drawn into a waste bottle. The vials then can be filled to a specified level with new buffer from the reservoir bottle. During subsequent activities the refilled vials can be used as inlet or outlet vials, or as flush and waste vials.

Preconditioning consists of a set of commands to specify a treatment of the inner surface of the separation capillary and to fill it with a reproducible composition of running buffer. A list of up to 100 entries can be specified with actions like "flush with pressure," "apply high voltage," "wait" a specified time, or "change" inlet or outlet vials to clean the electrodes of residual liquid.

During the injection a liquid amount of typically five nanoliters is placed into one end of the capillary. This can be done by applying high voltage for a certain amount of time or by forcing a pressure drop across the length of the capillary. While the voltage mode forms a plug flow profile, the higher-mobility species are injected preferentially, which has to be considered when quantitative measurements are needed. The pressure mode does not change the quantitative composition, but the hydrodynamic flow profile reduces separation efficiency. For more complex injection tasks a

table of serial actions can be programmed, for example a sample plug followed by buffer liquid.

During run time the electric field is applied to the capillary, which causes the different species to separate. The result is measured by monitoring the optical absorbance of the eluting liquid at up to five different wavelengths. For more complex tasks a full set of optical spectra can be acquired. For defect tracking and further evaluation, up to five other signals can be stored during the analysis. A timetable can be programmed to change parameter settings during the run in synchronism with the sample plug injection. Electric field, differential pressure across the capillary, and temperature in the air around it can be changed as easily as changing vials on both ends of the capillary.

After the measurement the separation path (capillary) can be cleaned by using a list of the same actions that were used for preconditioning. The capillary can be cleaned and filled with new liquid before the instrument is parked.

The final result of a measurement is a report that states the composition of the sample, both qualitatively and quantitatively. Using the spectral information, peak recognition and peak purity checks are possible. Calibration with previous analyses allows calculation of exact amounts. An experienced user can use the printout of an electropherogram to interpret the results.

Control Requirements

The electrophoretic separation works like an impulse response. The injected sample plug is the impulse introduced into the system. Because each component of the sample has a different specific mobility, the response is a series of (hopefully) separated impulses with different peak shapes: the electropherogram. This impulse response is dependent on both the input impulse (the injection plug) and the system response.

While the injection plug is mainly dependent on physical parameters, the system response is mainly dependent on chemical parameters, and both are dependent on timing relations. The injection plug depends on the geometry and length of the capillary, the viscosity and temperature of the liquid, and the value of the applied force multiplied by the time interval. While the first factors are not controlled but are more or less constant, the applied force and the time of application require good precision. If an injection of 50 mbar times 3 seconds is wanted with a precision of 1%, then a 50-mbar pressure source is required with 0.25-mbar precision and the time has to be controlled within less than 15 ms with transients taken into consideration.

The system response acts in two different ways. One is the bulk speed of the liquid (electroosmotic flow) and the other is the specific mobility of the separated compounds or molecules. These are primarily dependent on the length of the capillary and the voltage applied across it during the separation (during the run time), but are also dependent on the type of buffer and the temperature. There are also secondary dependencies, more chemical with the electroosmotic flow and more electrical with the conductivity and field distribution.

To address the electroosmotic flow dependencies the system allows activities bundled as "preconditioning." This is a series of up to 100 steps specified by the user to pretreat the capillary's inner wall. This treatment can include:

- Filling the separation capillary with mobile phase
- Removing gas bubbles
- Cleaning the stationary phase using extreme pH
- Exchanging the mobile phase
- Conditioning the stationary phase to the final analysis pH
- Reaching thermal equilibrium
- Removing old sample from the capillary.

While the freedom to select a number of different actions in serial combination allows the user to address various problems, the execution is reproducible and synchronized within 100 ms. This is valuable because the time constant to reach equilibrium may be days. To speed up the analysis, it is useful not to wait until equilibrium is reached, but to gain reproducibility by precision in timing. Thus, the results ride on a wave. In HPLC there is no specific timing requirement for other actions outside the run time. Different durations of injection cycles don't show a dramatic influence on the separation results. But in HPCE preconditioning is a major aspect. Reproducible results can only be achieved when the timing of the conditioning phase for repeated runs using the same method is always the same. It may be useful to have a constant time for feeding vials to the cassette, independent of the number of vials or the move distance.

Because communication from the PC-based software to the instrument via the HP-IB is too slow, the instrument must have all of the parameters needed before the analysis is begun. Therefore, the complete method is copied to the instrument, which then executes it completely without interaction with the computer.

Method Execution by Firmware

The analysis consists of user-definable steps like replenishment of vials, preconditioning of the capillary, injection, running, and postconditioning of the capillary. To achieve a high reproducibility of the analysis it is essential that the durations of the steps and the intervals between them have only small variations (< 100 ms). This is accomplished by generating the method on the PC and then downloading the method to the instrument, where the real-time operating system in the instrument takes care of the exact timing of the analysis. After the analysis the PC is responsible for data evaluation. The operating system in the instrument is a real-time, multitasking operating system previously used for the HP 1050 HPLC system.[2]

To address the conductivity and field distribution dependencies mentioned above the system allows selection of the control mode (voltage, current, or power) and the storage of the measured values of voltage, current, and power as raw data with the detector results. While voltage control is most precise when the conductivity is equally distributed across the capillary length, current control keeps the field strength in the buffer portion independent of the conductivity of the injection portion. If temperature is the major influence, power control can keep the internal temperature rise in the capillary constant to adapt to changes in overall conductivity. The stored measured voltage, current, and power data can be used for checking and diagnosis or to calculate specific behavior. For example, the mobility report uses the

voltage trace acquired during the analysis to calculate the apparent mobility of the species.

Temperature

The temperature of the capillary in the cassette has to be stabilized because it influences the electrophoretic results. One can compare this with the thermostabilization of an HPLC column. Electromobilities depend on solvent viscosity, which changes at a rate of about 2%/°C for typical buffer solvents. Chemical retention, if it occurs, is also influenced by temperature. In general, cooling is required because the current through the capillary heats the solvent and would boil the liquid with the sample in it. The temperature must be stable within at least ±0.1°C and perhaps as low as 10°C below ambient.

For cassette cooling a Peltier element is used instead of a chiller. This element is made of semiconductor material, and with electrical current applied, one end is heated while the other end is cooled. The direction of current flow defines which end is cooled. Thus the unit can be used for both heating and cooling by reversing the excitation voltage.

During automated sequences the samples to be analyzed are kept in the liquid handling tray. If temperature-sensitive samples like biosamples are being analyzed, the sample storage area has to be cooled to as low as 5°C to ensure that the samples are not spoiled while they are stored in the instrument. While capillary cooling is always required, the tray cooling is an optional feature. The user can connect a chiller to control the tray temperature and the HP CE instrument will measure and display the tray's temperature.

Forced-air cooling is used. The forced-air circulation has to be fast enough to ensure that water condenses only on the surface of the cooling heat exchanger. This is important because a wetted surface is able to conduct electrical current. With 30 kV in a sample bottle this can lead to unexpected high voltages in other areas of the instrument.

An analog-to-digital converter (ADC) resolution of 12 bits is sufficient for all of the signals that need to be converted. The highest resolution is required for temperature control of the capillary and for raw data generation. A 12-bit ADC allows a resolution of 0.075 μA in current and 0.025°C in temperature measurements.

One ADC is multiplexed to measure the analog signals for voltage, current, return current, pressures, and temperature. Multiplexing requires more firmware support. It is a load to the interrupt level of the system processor and is therefore time-sensitive, but it saves hardware cost.

Separation Subunit

Fig. 1 shows the functional blocks of the separation subunit, which are:

- Operating system
- Vial and liquid handling
- High-voltage control and safety functions
- Thermal control
- Replenishment and injection.

These functions are separated onto different printed circuit boards. The separation subunit contains the following boards (see Fig. 2):

- Main processor board (CMP)
- PC interface board (CRB)
- Electromechanical interface board (EMI)

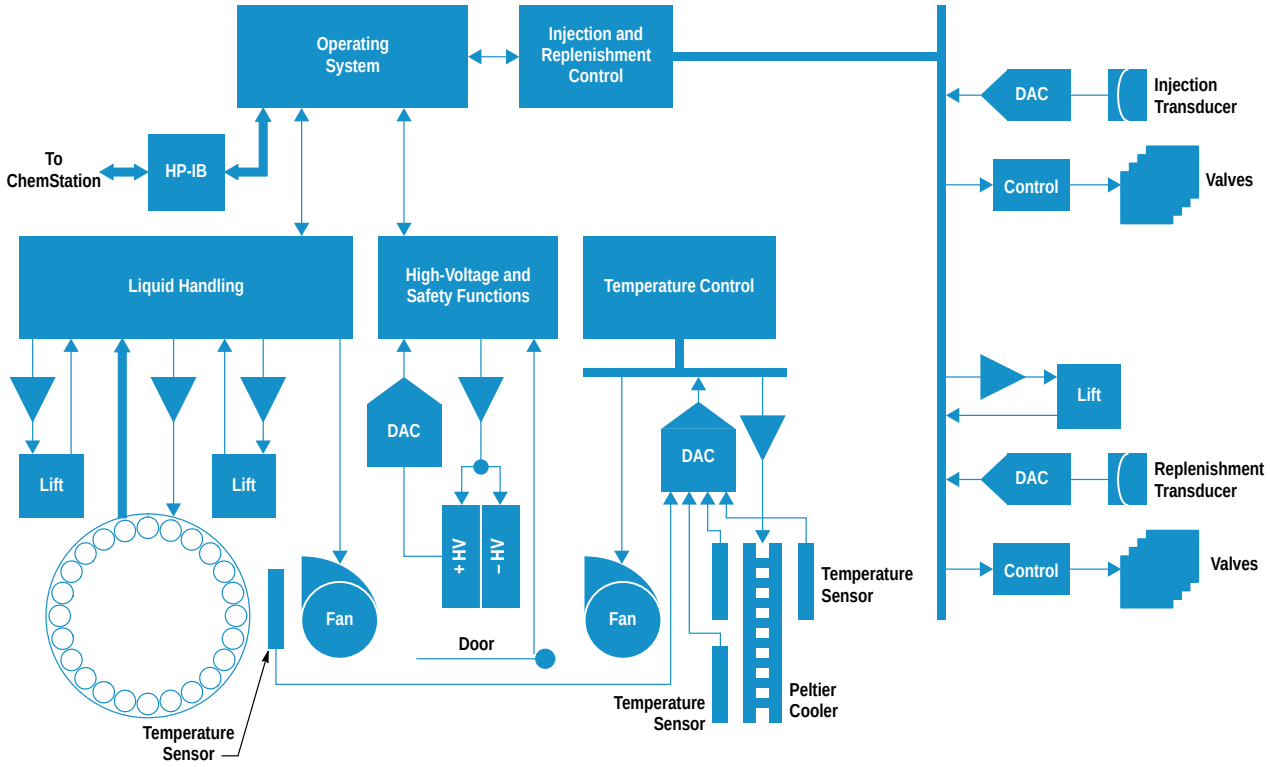


Fig. 1. Functional block diagram of the separation subunit.

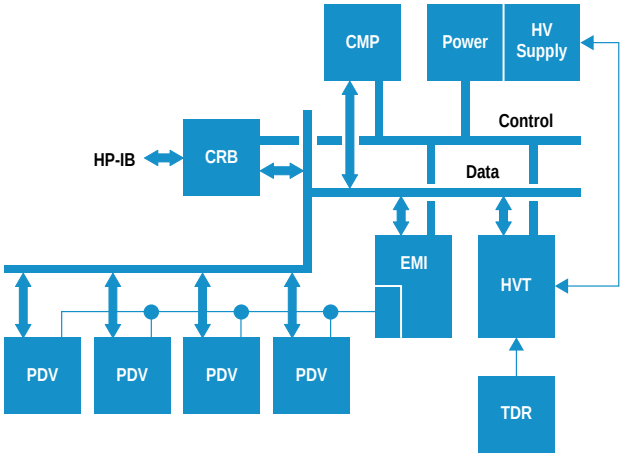


Fig. 2. Electronic block diagram of the separation subunit. CRB = PC interface board. CMP = main processor board. EMI = electromechanical interface board. PDV = servo boards. HVT = high-voltage and temperature control board. TDR = temperature driver board.

- Four servo boards (PDV)
- High voltage and temperature control board (HVT)
- Temperature driver board (TDR) for Peltier current control.

Position Control Servos

The central functionality of the separation subunit is vial handling. Up to 50 vials can be selected to be placed at either the inlet or the outlet end of the separation capillary. In addition to taking a vial from the tray and placing it at a capillary end, the vial handling system has to place the vials at a defined height. Depending on the liquid level in the vial and the necessary depth of immersion for the electrode, a positioning offset can be set. A position servo system is used to drive the lift spindle, which is able to provide position and force feedback.

The same type of position servo is used for the tray movement. The servos are initialized to a special position and all vial positions are in reference to that. For the tray this special position is a fixed distance from the mechanical stop that prevents the tray from being turned more than one revolution.

Every movement using the servo positioning is controlled by a smooth harmonic function. Both the speed and the acceleration are ramped. This generates a stress-free moving profile that is easily calculated by the microprocessor at an 80-Hz interrupt rate.

Tray and Lift Initialization

After power-up of the instrument the lifts and the tray must be initialized to a known state before the firmware can use them. The initialization procedure first interprets the current state of the mechanics, then finds the home positions for the servos, and then calibrates the lift positions relative to the segments of the tray. The sequence allows precise positioning without adjustments of the mechanical assembly. The need for absolute position sensors is avoided by using the position and torque information from the servo drivers. This lowers the cost, since even inexpensive sensors such as light switches and Hall sensors cost money and require cabling, connectors, and often adjustments. Furthermore, the presence of sensors in the high-voltage area could lead to reliability problems.

The mechanism is designed to allow random access by the lifts to the vials in the tray. To achieve this the lift finger is lifted through slits in the tray segments and the movement of the tray is blocked when the lift finger is within a slit. This means that the system needs to know the exact position of the tray segments relative to all lifts. All servos need absolute reference points for proper positioning. To bring the vials in the lift stations back to their correct locations after a power failure, the location of the vials is held in continuous memory.

To find the current state of the system the initialization procedure first tries to initialize the tray. In most cases this is successful because the lifts have normally reached their upper or lower positions before the instrument is switched off. If the tray is blocked by one or more lifts the next step is to resolve the blockage. The lifts are moved down with reduced torque. Because the orientation of the tray relative to the lifts is unknown in this phase, it can happen that the lift finger is blocked by a vial or by a tray segment. In this case the lift is moved upward again out of the tray. Discrimination between hitting a vial or a segment and a correct initialization of a lift is based on the characteristics of the torque increase when the lift finger hits something hard. The real endstop is much stiffer, so the torque increase as a function of distance is steeper compared to the situation when the lift hits a vial, for example. After this step no lift should be blocking the tray.

Now the initialization and find home process starts. First the tray starts to initialize again, checking that its complete path is free. It does a full turn while monitoring the force to detect blockages. The lifts are now initialized and the vials in the lifts are brought into the corresponding tray locations. If there is any vial conflict, for example the user put a vial in location 20 of the tray while a vial with the same number is already in a lift station, the system sends an error message to the user to remove this vial first, and the instrument turns the tray so that the corresponding location moves to the front access region. Because of problems like vial conflicts, the initialization process needs some user interaction. It is not recommended to start operation automatically at turn-on of the instrument.

The next step is to calibrate the position of the lifts relative to the tray. Mechanical tolerances in mounting the lifts and segments are compensated by this calibration procedure. The lift finger of a lift is brought into the slit of segment 0 and the tray is moved with constant torque to the left and to the right. The average of the positions reached gives the exact tray position for that lift. With the same procedure the position of every tray segment relative to the lifts is measured and stored in the continuous RAM. Every access to a vial uses the stored values to eliminate mechanical tolerances.

Servo Control

An application-specific servo chip, optimized for velocity control, is used in this instrument. Special firmware routines generate the velocity profile to achieve a gentle movement of the servos. The servo chip uses current control to drive

the motor, and the control error can be used as a signal proportional to the required servo torque. This allows full control over the forces in the system. It also makes it possible to detect that a servo is blocked and switch to the error recovery routine to resolve the blockage. To achieve a good dynamic response of the servo, the motor driver can generate a maximum current that will destroy the motor with just eight seconds of continuous operation. Therefore, to avoid damage we implemented a firmware fuse. The firmware calculates a simulated temperature of the motor according to the heat generated by the current through the winding, subtracting expected thermal losses. If this simulated temperature becomes too high the servo is switched off and an error message is generated.

Movement Control

The movement control firmware translates high-level move instructions into low-level move commands for the servos and a signal to indicate when an instruction has been executed. The high-level instructions include:

- Find the home position for a servo
- Remove the vial from a specified lift
- Find a vial, put it in the lift, and raise it to a specified position
- Set torque (move a servo with a specified torque)
- Put the two home vials in their lifts and raise them to their normal positions.

The low-level move commands include:

- restartservo. Find the orientation of the rotor and stator for a servo.
- initservo. Find the zero position and search index.
- segmentcheck. Measure the position of a tray segment relative to a lift.
- move. Move a servo to a given position with a velocity profile.
- findvialpos. Find the location of a specified vial.
- findvialinlift. Find which vial, if any, is in a specified lift.
- storevialpos. Store the location of a vial in continuous RAM.
- Signal the end of the high-level command.

For example, say that vial 30 is in the replenish lift and vial 25 is in the inlet lift and the task is composed of the high-level commands, "Find vial 30, put it in the inlet lift, and raise it to the default position." This is translated into the following low-level command sequence:

- findvialpos of vial 30. This will find vial 30 in the replenish lift.
- move vial position 30 of the tray under the replenish lift.
- move the replenish lift downwards until the vial is in the tray.
- storevialpos of vial 30 in the tray.
- findvialinlift in the inlet lift. This will find vial 25 in the inlet lift.
- move vial position 25 of the tray under the inlet lift.
- move the inlet lift downwards until the vial is in the tray.
- storevialpos of vial 25 in the tray.
- move vial position 30 of the tray under the inlet lift.
- storevialpos of vial 30 in the inlet lift.
- move the inlet lift upwards until the default position is reached.
- Signal the end of the command.

These low-level move commands are put in a queue. The servo control firmware executes one element after the other from this queue. When a new high-level command arrives before the queue is empty the routine that translates the high-level commands to low-level commands can't use the actual vial locations because the commands that are still in the queue may alter these locations. To follow the movement of the vials, every entry in the queue creates a virtual vial location which represents the vial location after this entry is executed. When a movement error occurs the queue is cleared. Depending on the type of error, the error recovery routine generates a new queue to bring the mechanism into a controlled state and gives a description of what caused the error.

High-Voltage Control and Analog Signals

The high-voltage power source is an encapsulated unit that contains two cascades, one for each polarity. A high-voltage vacuum relay is used to switch polarity according to the method settings. To interface the high-voltage power source to the microprocessor, some hardware is used to generate analog control signals and some monitoring signals are fed to the ADC. As mentioned above, the ADC is multiplexed to measure the analog feedback from the high-voltage power source, the pressure and vacuum signals, and the temperatures. Digital status information from the high-voltage source is fed to the microprocessor for control and error detection.

A double-contact relay is implemented for safety purposes. In any emergency condition this relay breaks the 24V supply for the high-voltage power source. The double contact (two contacts in series) ensures the required safety level for this type of functionality.

Voltage, Current, or Power Control

Historically, voltage control has been used for electrophoresis. This keeps the field strength constant during the analysis. However, even the best cooling and thermostating of the capillary's outer wall cannot prevent a temperature rise inside in the liquid, especially when using a high salt concentration. A temperature rise changes the viscosity and thus the mobility of the various species. This makes the different peaks elute faster, even when the field strength is kept constant. Because the conductivity of the buffer liquid also increases with rising temperature and thus the joule heating is increased, this effect is self-amplifying.

In such a situation, current control is advantageous. With current control, the feedback loop causes the voltage across the capillary to decrease when the conductivity increases as a result of an internal temperature rise. Thus, the effect is more self-limiting. However, measuring very small currents at high voltage is difficult, so the precision of current control is not as good as voltage control: 1% compared to 0.1%.

In thermally sensitive applications it may be advantageous to use power control. Variations in conductivity are then compensated, leading to a more reproducible internal temperature, which in turn stabilizes the analysis results.

Voltage Measurement

The high-voltage supply outputs a signal proportional to the high voltage at the output. This signal is used for monitoring, power control, electromigration injection, raw data to the PC for data evaluation, and limit sensing.

The output voltage is controlled by the high-voltage source internally. This means that voltage measurement is not required for voltage control. The major performance requirement is the precision and stability of the output voltage; absolute accuracy is secondary. The HP CE design yields a precision of 0.1%, which means only 10V fluctuations on a 10-kV output.

Current Measurements

Two current measurements are made on the high-voltage source. One is the current monitor signal from the high-voltage power supply and the other is the current at the capillary outlet next to the detector. For the latter, there is a current sense resistor between the vial ground electrode and the power supply's grounding pin. This leads to the requirement that the ground electrode be insulated. Even when there is no voltage difference, the current must be confined until it is measured. This measurement detects current from the high-voltage supply that is not flowing through the capillary. Such leakage current can be caused by wet surfaces or carbon tracks in plastic parts. It is a good safety and diagnostic feature to detect such leakage traces in the high-voltage path.

The measured current data is used for monitoring, limit sensing, power control, electromigration injection, raw data to the PC for data evaluation, leakage current detection, and bubble detection. For generation of raw current data only the return current from the capillary is used. This signal is integrated to determine injection volumes during electromigration injection when the current and time have been specified.

Corona effects, sparking, and insulation breakdown are detected and a yellow warning flag is displayed on the PC user interface to point out these problems.

To stop the analysis when a gas bubble has formed inside the capillary, a lower current limit can be set. If the current is not over this limit half a minute after voltage is applied, the analysis will be terminated to prevent local burning of the capillary.

Pressure Measurements

Data from pressure measurements is used for control of the pressure source, hydrodynamic injection, liquid level sensing during refill, raw data to the PC, and diagnosis and error detection.

For various needs, the pressure source can supply up to 1 bar of pressure and 0.4 bar of vacuum. Both pressures are stored in reservoirs which are filled sequentially from a membrane pump. This bipolar source allows push-pull actions which are used for empty-refill operations, up-down injection profiles, or controlled bipolar pressure across the capillary.

Injection Control Algorithm

To get high reproducibility for quantitative measurements in CE it is essential to have a very precise injection process. The HP CE instrument allows either hydrodynamic or electromigration injection. In the hydrodynamic injection mode the injection volume is determined by the time integral of pressure across the capillary. This pressure/time integral can be as low as 20 mbar·s. When pressure is applied by simply opening a valve for a specified time the slope of the pressure during switching of the valve is not very reproducible. Gas rushing through the system will cause pressure drops that may result in transient errors in the pressure sensor signal relative to the real pressure generated across the capillary. Therefore, as described earlier, the HP CE instrument uses pressure tanks as gas flow sources with high output impedance. The conditions are controlled so that the pressure increases or decreases according to a linear ramp function. With the bipolar gas sources this ramping can be performed in both directions. With this system, valve switching time, the resistance and capacitance of the tube and chambers, the signal sampling rate, and even limited leakage are no longer factors that influence the precision of the pressure integral applied across the capillary.

To finish the injection profile another valve is switched to ramp down the pressure in the inlet vial. A kind of pulse width modulation is used to achieve a predefined slope. At specific pressure values during the ramp the system checks to see if the integral has reached the appropriate value. By design, the integral will always be smaller than the expected value. The ramp is then stopped and the pressure is held passively until the measured integral is correct, and then the ramp starts again. This leads to a successive approximation of the integral until the target value is reached within ±0.3%.

For hydrodynamic injection a variable pressure can be selected in a range of ±50 mbar. For flush and clean procedures the full pressure of about 1 bar is applied to the inlet vial.

Replenishment Control

During the analysis the content of the run buffer will become modified by the addition of electrochemical by-products. To achieve good reproducibility of the analysis the run buffer should be exchanged after several runs. The replenishment system allows the user to empty a vial, refill a vial, or bring a vial to a desired level automatically.

To determine the liquid level in a vial the following steps are executed:

- Draw a small airflow through the short end of the replenish double needle and measure the pressure drop in the needle.
- Move the vial slowly upward against the end of the capillary and wait for a sharp increase in the pressure drop, indicating that the needle has touched the liquid surface.

The noise of the pressure sensor must be very small because the detection algorithm uses the slope of the pressure to achieve a small delay between touching the surface and detection. For a complete description of the replenishment system, see the article on page 32.

Capillary Temperature Control

Because a temperature change of 1°C changes the viscosity by more than 2% and viscosity has an influence on electrophoretic mobility, the capillary temperature must be controlled so that it is stable within ±0.1°C so as not to degrade the analytical reproducibility.

The capillary temperature is controlled by heating or cooling, depending on its relation to ambient. A Peltier cooler is used as a cooling element for the capillary. By reversing the current polarity the same element can be used for heating. Because its power efficiency may be as low as 20%, an electrical supply of 100 watts is required. A current source was selected to drive the Peltier element because in this type of

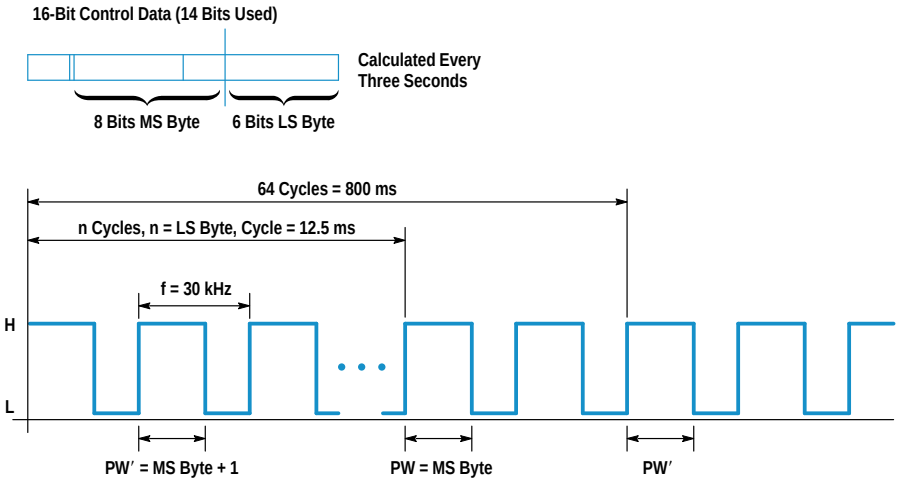


Fig. 3. A six-bit fine-tune function is added to the basic 8-bit pulse width modulation of the switched-mode current source that drives the Peltier cooler. This increases the precision of the capillary temperature control.

application the Peltier is much more linear than with voltage control. An additional benefit is that current control is independent of the supply voltage, which allows operation with an unstabilized voltage, reducing the cost of the power supply and increasing the overall efficiency.

The efficiency of a Peltier cooler is dependent on the temperature difference between the hot side and the cold side. At a certain temperature difference there is no more cooling even when more electrical power is supplied. To avoid operation in this region both Peltier side temperatures are measured. This information is also used to prevent either side of the Peltier from becoming heated above 80°C, which would decrease the lifetime dramatically.

The temperature to be controlled is not the Peltier's, but the capillary's. While the exact temperature of the capillary is hard to measure because of the size and geometry, the air temperature around the capillary tracks it closely in value and time. Therefore, a sensor is placed where the airflow exits the capillary cartridge. This allows a fixed location for the sensor and wiring while the cartridge is easy to access. The high airflow speed of 10 m/s provides good, fast thermal contact between the capillary and the thermosensor.

Cascaded Control

To achieve good precision combined with steep slopes, a cascaded control concept is implemented. This requires a little more hardware and firmware, but offers better controllability along with flexibility and ease of optimization. This saved risk and speeded the development process.

As already mentioned, the efficiency of the Peltier is strongly dependent on the temperature difference between its hot and cold sides. Therefore, it is wise to keep the temperature difference between the air and the Peltier as low as possible. This leads to a heat sink large in size and volume. A good heat sink material is metal, which has a high thermal capacitance. The result is a long dead time and a high time constant between the Peltier temperature and the air temperature.

Conventional control algorithms are either slow or show long ringing or instabilities. The cascaded control loop of the HP CE instrument uses a sensor close to the Peltier's cold side to achieve speedy measurement for steep slopes and good stability. However, the ambient rejection is not so good. Varying loads, dominated by the ambient temperature, may lead to a temperature difference of up to 8°C between the heat sink close to the Peltier and the air at the capillary.

To correct this problem, a second control loop is implemented around the control loop that stabilizes the temperature at the Peltier. This outer loop overrides the setpoint value for the first loop and adapts to ambient changes in the "minute" range while the inner loop of the cascade operates in the "second" range. As soon as the inner loop reaches a temperature close to its setpoint, the outer loop is activated to adapt the setpoint accordingly to reduce the difference between the capillary air temperature and the user-programmed value.

The Peltier element is mounted on a heat sink built into the top rear area of the instrument above the two card cages of the subunits. The switched-mode current source is on an extra printed circuit board next to the Peltier. The power transistors are connected to the same heat sink as the Peltier's hot side and therefore can be cooled efficiently. The current source is controlled by two digital signals. One signal controls the power direction and the other defines the current through the Peltier.

The pulse width modulation for the current source has a frequency of about 30 kHz with 180-step pulse width resolution using a clock of 6.66 MHz. An 8-bit data word is needed for this, but if only 8 bits were used, the limited resolution would require a fast response to keep the resulting fluctuations small. Therefore, a 6-bit fine-tune feature first used in the HP 1050 HPLC system is used for Peltier control to keep the temperature stable within ±0.1°C or better. This gives a controllable dynamic range of about 10000:1 with the highest resolution at low power values. The control loop calculates a 14-bit control signal every three seconds. This calculation is somewhat complicated, but is not executed very often. The upper 8 bits are fed to the pulse width modulator, while the lower 6 bits are used to modulate the pulse width modulator in a 12.5-ms time slice (see Fig. 3). The benefit is a higher base frequency, which is filtered better by the thermal time constant of the system. Thus, the required temperature stability is achieved with inexpensive digital hardware, simple firmware algorithms executing at an 80-Hz rate, and complex calculations every three seconds.

Safety

To guarantee user safety, the HP CE instrument implements a number of safety features. Tray rotation is stopped after a maximum of 400 ms when the tray door is opened. Any movement that was in progress is finished after the door is closed again. The high-voltage power supply is brought to 0V and switched off when the top cover is opened or when a leak is detected. The cassette fan is switched off when the top cover is opened to avoid finger injuries. Pressure on the vials is released immediately when the top cover is opened; otherwise, the vial contents would be sprayed inside the detector compartment. Pressure is removed from the buffer bottle if a leak is detected.

Evaluation

When a concentration dependent detector signal is integrated over time the peak area is a measure of the amount of separated sample compound. The sample area and the amount are related by the speed of the absorbing particles as they move through the detector cell.

In HPLC separations the speed of the sample is normally known. It is the flow rate of the liquid. In electrophoresis, the separating power is the difference in the mobility of the compounds, which means that the compounds have different specific speeds while moving through the detector.

To correct for speed differences in later evaluation steps, information about the moving force is needed. In a simple analysis the voltage is constant during the complete run, but temperature changes have an influence on viscosity and mobility, and a time-programmed voltage or fraction collection can have a tremendous influence on absolute-area-to-amount calculations.

To support future evaluation developments the HP CE instrument supports five raw data channels in addition to the detector signals:

- Voltage, current, and power of the high-voltage source
- Pressure across the separation capillary
- Temperature of the air from the cassette.

Diagnostics

Traditionally, diagnostic routines built into instruments are based on the perception that the failure modes are already known during the development phase. For these failure mechanisms special tests are implemented (RAM, ROM, communication, display, etc.) which generate a pass/fail indication or a value. Sometimes the knowledge of a specific failure mode in the early development phase leads to changes of the product that reduce or even eliminate the failure mode. However, there is no possibility of avoiding the learning curve. Unexpected behavior and dependencies always appear after the product is released, and may increase the list of diagnostic functions needed. Thus, the diagnostics implemented in the early phase might be of little use while those most needed are not supported. The disadvantage of the traditional concept is that the diagnosis of failure modes discovered after the development phase is not possible without product changes.

The HP CE instrument implements traditional diagnostic functions as outlined above, but the firmware also implements basic functions for diagnosis, such as read sensors, or write to hardware address. Macros on the PC can use these basic functions to construct complex diagnosis functions to address specific problems. Now if a new failure mode is discovered, we just add a diagnosis macro for it and create an updated version of the diagnostics disc to distribute it to the manufacturing and service organizations. This leads to a continuous improvement of the diagnostics for the instrument, and we spend the time to create diagnosis functions only for failures that have actually occurred in the field. A valuable side effect is that most module test routines are implemented within the product, so manufacturing does not need a large amount of special equipment. The service person at the customer's site has all the equipment needed to perform the same tests as originally used in the factory.

Diagnosis Functions

The instrument normally collects equally spaced data for voltage, current, power, vial pressure, and temperature in a buffer memory in the instrument RAM. The diagnosis firmware makes it possible to redirect other data like servo speed, servo position, servo current, sensor data, valve states, and so on to this buffer. Thus, we can store up to five configurable measured values, with a configurable data rate of up to 20 Hz. The PC software reads this buffer and generates reports and graphics.

This tool greatly improved our ability to find bugs during the development phase because we got the history of servo positions, pressure values, and other measured values when an error occurred. It also gave us much more insight (like a built-in oscilloscope for electromechanics) into what really happened in the unit than we would have obtained from single measurements.

An advantage of this firmware feature is that service personnel can use it to find problems in the field and can send the printout to the factory for a simple form of remote diagnosis. An example is the macro for checking the pressure system. It takes about five minutes and checks the power of the air pump and the tightness of the pressure and vacuum system, the injection system, and the replenish system and stores the data from all of the pressure sensors. A graph of these quantities characterizes most of the pressure system and helps diagnose the problem if there are deviations from the normal behavior.

Acknowledgments

The success of this product is a result of the exceptionally good teamwork of all the people who worked on it. We especially want to thank Gerda Biselli for her patience in writing the online help and the manual, Arno Graf for the discussions about serviceability and diagnostics, and Claudia Maschke for designing the temperature controller board. The great support of the production and marketing groups during development helped us to find and realize some of the ideas that make the instrument as good as it is.

References

1. A. Wiese, et al, "A New Generation of LC Absorbance Detectors," *Hewlett-Packard Journal,* Vol. 41, no. 2, April 1990, pp. 36-43.
2. C. Büttner, et al, "Firmware Development for a Modular Liquid Chromatography System," *ibid,* pp. 44-50.

A User Interface for Capillary Electrophoresis

The graphical user interface of the HP CE instrument is designed to be easily understood by users familiar with other separation methods but new to CE. It provides for method programming and simulation and for visualization of the status of the instrument and the running analysis.

by Alwin Ritzmann and Klaus Witt

The architecture of the user interface for the HP CE instrument is similar to HP's LC ChemStation. Like the LC ChemStation, it uses Microsoft® Windows as a platform. Because electrophoresis is seen as similar to HPLC and the same type of diode array detector is used, it made sense to us that the user interface should be similar. It also helps that both systems are developed at the HP Waldbronn Analytical Division.

The HP CE ChemStation consists of an HP Vectra 486/66XM or compatible computer with 12M bytes of RAM, a 420-Mbyte hard disk drive, an HP-IB (IEEE 488, IEC 625) card, MS-DOS® 6.xx, and MS Windows 3.1 or higher. The PC is in charge of data handling, which includes all functions with midrange timing requirements:

- Sequencing
- Collection of method parameters
- Method storage
- Long-term control of the instrument
- Error handling
- Collection of run-time data
- Data processing (e.g., integration)
- Report generation
- Message logbook and error log
- User-controlled liquid movements.

User Interface Strategy

A major objective for the user interface was to help the user not be afraid of the measuring device. Because CE is a new separation technique, many users will be switching from other procedures they are already familiar with, and their expectations may be basically different. A person acquainted with slab gel electrophoresis will have no problems with high voltage, but sequence analysis programming and automatic data analysis with calibration will be new. A chromatographic

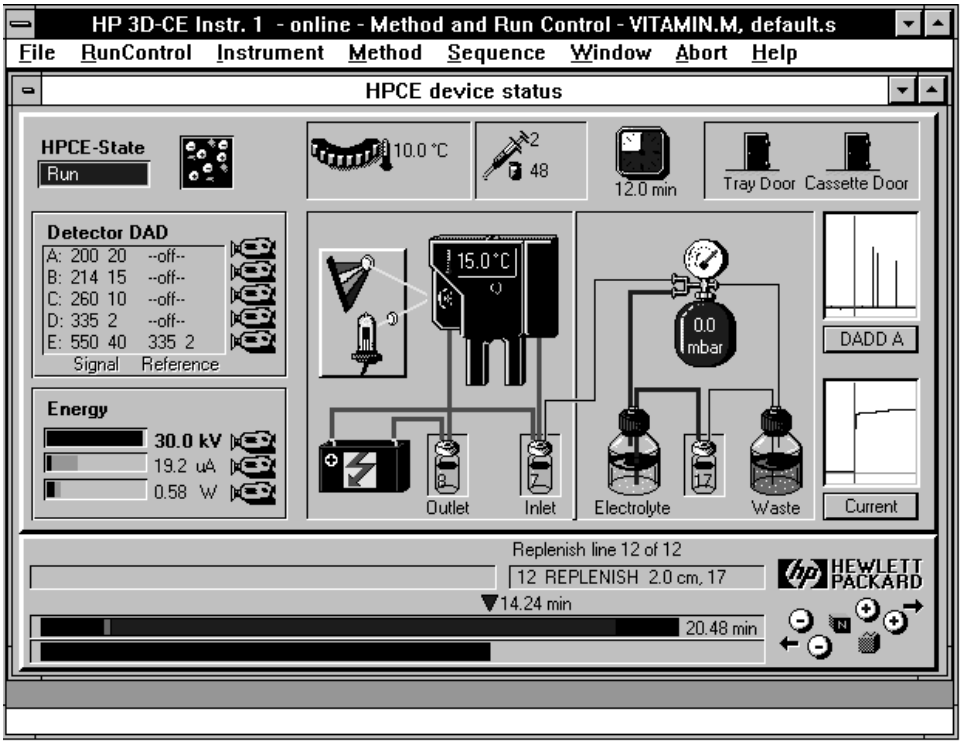


Fig. 1. Top-level window of the user interface of the HP CE instrument.

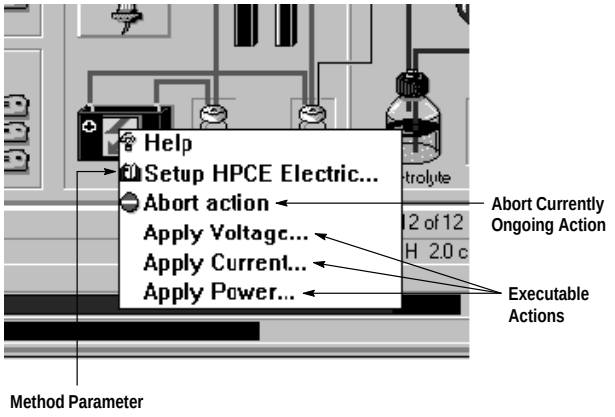


Fig. 2. A pop-up menu appears when a symbol is selected by a mouse click.

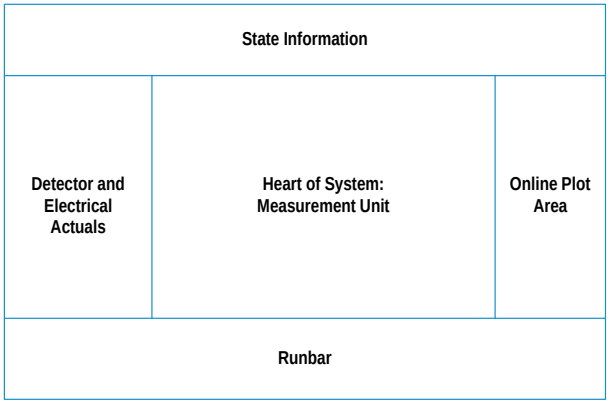


Fig. 3. Layout of the top-level window.

expert, on the other hand, will consider these aspects old hat, but step-by-step conditioning and the frightfully high voltage may be new.

A graphical user interface was selected to provide visibility into the operation of the instrument and thereby counteract any initial problems by conveying the feeling to the user that the measuring device is controllable. Communication with the user by interactive visibility is a great advantage. Many applications cannot be imagined without icons, 3D buttons, WYSIWIG, or multimedia anymore. It is far more comfortable to follow the consequences of one's behavior on the screen interactively.

The graphical user interface of the HP CE system is designed particularly to address the following three objectives:

- Visualization of a programmable but automatically running process, as feedback to the user
- Context dependent or function-related access to specific settings or value groups
- Manual control of specific functions for which direct triggering is impossible or unreasonable in automatic operation.

Visualization means not only the extensive but clear representation of as many values, activities, or processes as possible, but also the ability to simulate a programmed method.

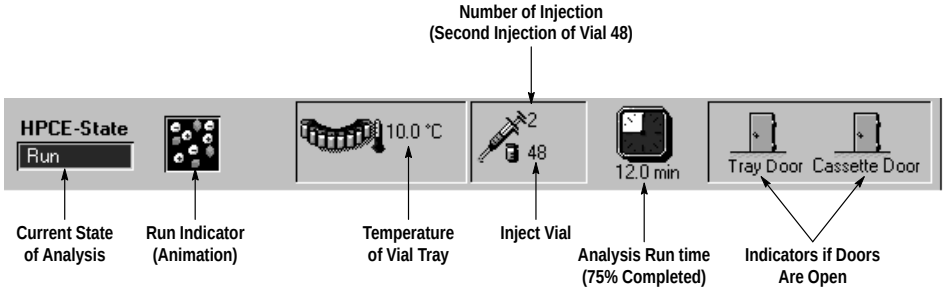


Fig. 4. Status line.

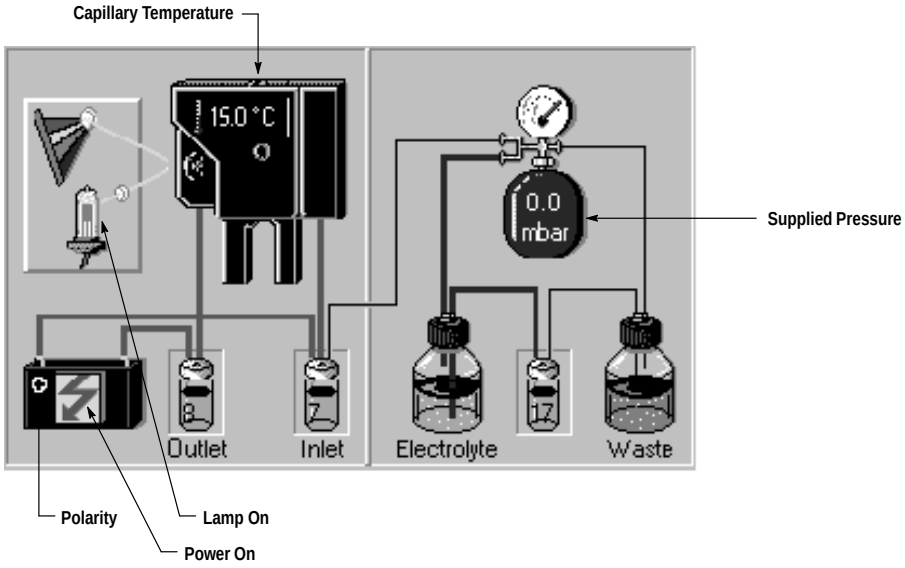


Fig. 5. Measurement unit display.

Development of a Common ChemStation Architecture

The HP CE user interface is based on a common HP ChemStation architecture. The architecture and many software components are shared with ChemStations previously developed for other analytical techniques. For example, because most of the HP CE ChemStation functionality is identical with the high-performance liquid chromatography (HPLC) ChemStation (automation, data acquisition, processing of measured data from the detector, and reporting), about 80% of the HPLC ChemStation code was reused. This article describes the evolution of this common software architecture and the core software components.

The term ChemStation was created to describe the use of a workstation as a tool for an analytical chemist. Independent of the analytical technique, the user works with a ChemStation to control the analysis system, acquire measured data, process the data, and create the analysis report. In addition, these complex functions must be fully automated for unattended operation. As PCs became very popular for the chemist's desk as well as in the analytical laboratory, HP product generation organizations, together with a system development group, defined the following objectives:

- Develop a ChemStation architecture based on the Microsoft® Windows platform.
- Develop reusable components from which various products can be assembled.
- Develop selected architecture-neutral components to be used for other data system platforms (e.g., UNIX® system-based workstations).

The architecture for the ChemStations is based on the available Windows technology. The development environment is based on the Microsoft development tools (C, C++, MS foundation classes, etc.). Both the Windows platform and the environment have evolved significantly during the last couple of years.

Because the common software components were developed by an international team, an appropriate infrastructure was set up to support the engineers working at two sites, one in the U.S.A. at Wilmington, Delaware and one in Waldbronn, Germany. A history management system (HMS), telephone conference equipment, small decision making groups, and system build processes were defined and implemented. The product development program was set up so that the software development groups were not isolated from the instrument development groups.

The infrastructure as it exists today is much different from the way it was defined in the beginning. As we learned how to deal with such a complicated environment, we went through several steps with the infrastructure and the organizational structure to make sure that the various software product generation groups were linked together.

The architecture and the development of the ChemStation products also went through several evolutionary steps. We always focused on clear product examples during the development of the core components. This approach made sure that the software development met the instrument requirements for the various analyical techniques. We also had to maintain the architecture and the core components according to the fast-changing Windows platform technology.

The first step was to introduce software products for HPLC, HPCE, and UV/Vis instruments. All these software products were built on the available common software components. As an example, the analytical language processor, a BASIC-like language, is used in all the products as a common key component. In a second step we went further and built one core software product supporting HPCE, GC, HPLC, and analog-to-digital converters. "Plug and play" modules specific to various instruments and analytical techniques can be added to the core software product. This is the basis for the recently introduced multitechnique ChemStation, which provides control, automation, and data handling for up to four GC, LC, or CE instruments and an optional analog-to-digital interface. This product is always delivered with the core software and an individual module for one specific analytical technique. If the customer wants to add another instrument, either for the same analytical technique or to combine techniques (for example, HPLC with CE), another module can be ordered and installed to work with the core.

This approach is very efficient and highly appreciated by our customers and field people. If the customer knows how to operate one product it is easy to operate the ChemStation for another analytical technique. The customer can concentrate on the analytical technique and always finds the same user interface. Other techniques, such as UV/Vis, SFC, and so on, share all the common components from the core and therefore are on the same architecture.

This approach has resulted in a very high reuse level—between 65% and 85%—and ensures that all the products meet high quality standards with a reasonable amount of testing and software validation effort.

Development of the common ChemStation architecture was only possible through the involvement and cooperation of many people throughout various HP organizations, who deserve a great deal of credit for these achievements.

Herbert Wiederoder
Section Manager
Waldbronn Analytical Division

Simulation is an advantage when, for example, a process that will run for a long time needs to be checked quickly for completeness and consistency, or when a method is being edited whose execution would not be possible given the present state of the measuring device.

User Interface

The HP CE ChemStation has one .EXE file and one top-level window, including the command processor. The top-level window contains child windows that are activated on user request. A status window and the online diode array detector plot window are shown by default. The top-level menu is user-definable and supports additional short menus.

As the name capillary electrophoresis indicates, we are dealing with an analytical measuring method taking place in a capillary with the help of an electrical voltage. The capillary is selected by the user based on specific criteria and is reeled up in a cassette which is then inserted in the measuring device. A picture representing the cassette forms the center of the top-level window of the graphical user interface (see Fig. 1).

In the top-level window, the state of the HP CE instrument is shown. Any important information is obvious at a glance. In accordance with the setup in the instrument and following the function as it is seen by the user, the respective functional

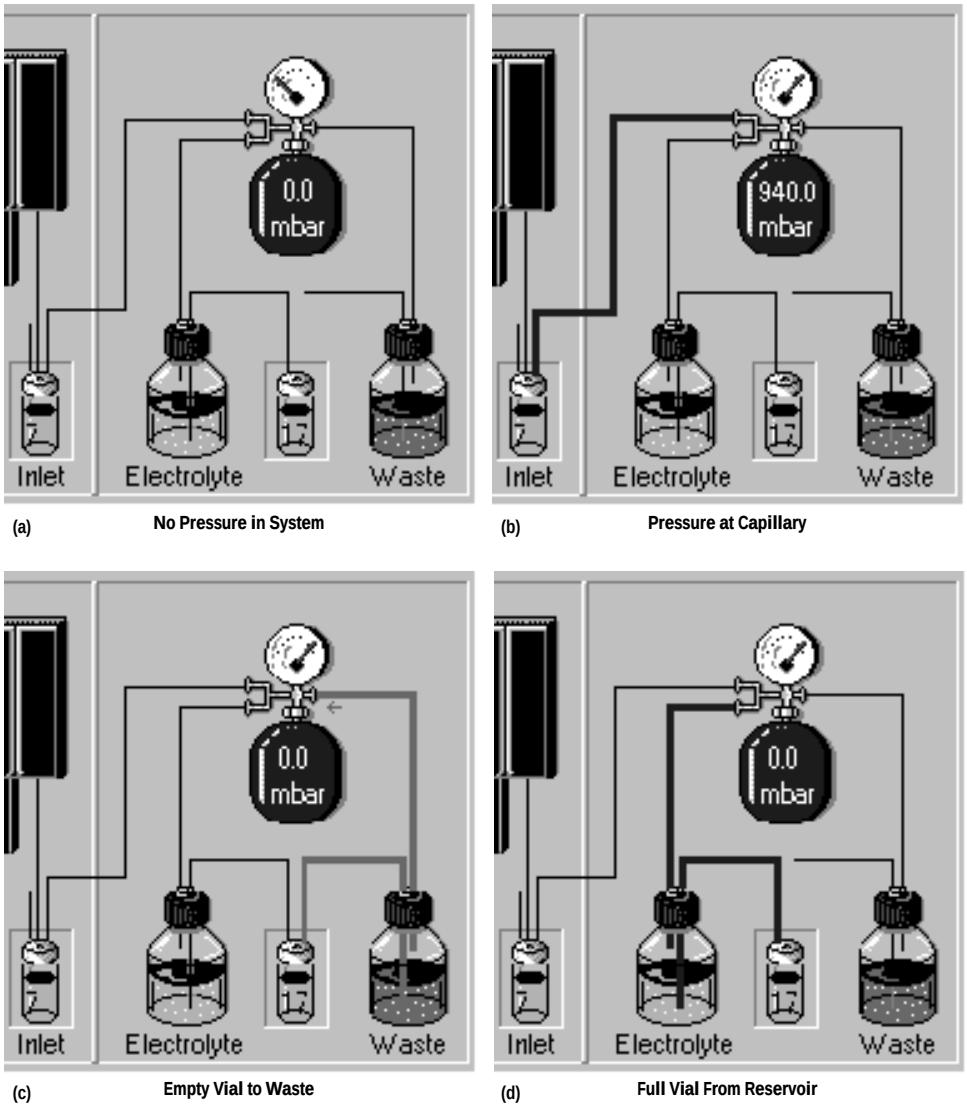


Fig. 6. Replenishment system displays. (a) No pressure in system. (b) Pressure at capillary. (c) Vial being emptied to waste. (d) Vial being filled from reservoir.

elements are arranged around the cassette. The user can immediately associate functional modes with the stylized pictures and thereby get a clear impression of the process and present status even without having much experience.

Each group of functions is represented by an icon or symbol showing the current state. A user who has a specific intention can select the component on the graphical interface that comes closest to it. If the mouse pointer is clicked on a component, a pop-up menu field appears close by, listing the functions linked to the component. In the case of the detector, for example, the pop-up menu offers lamp on/off or wavelength selection. When the battery icon is clicked, a pop-up menu of electrical parameters shows up (see Fig. 2). A specific icon in the menu field (the book symbol) immediately shows whether an action results in the modification of the method.

Fig. 3 shows the layout of the top-level window. Wherever possible, a combination of analog and digital display was chosen. Analog display facilitates the quick qualitative grasp of data by the human mind. Precise representation is realized by the additional digital display. In the analog display of the electrical parameters, for example, bar charts make states quickly visible to the user without too much thinking. For example, one can recognize at once whether the programmed value has been reached and how far away the actual value is from the system limit. These things could be conveyed with the digital display alone, but would be much harder to comprehend.

Status Line

In the left corner, the topmost line shows the current state of the instrument (Fig. 4). Different colors are used to help the user visualize the instrument state. For example, green is used for all ready, blue for running, yellow for not ready, and red for error states. Text matching the respective state is displayed in the field. To the right, there is a run indicator in which an animation starts running as soon as an analysis begins. The animation quickly provides optical feedback that an analysis is in process without requiring the checking of different states beforehand.

A clock in the status line displays the times for different actions. It is not a 24-hour clock, but displays the time elapsed relative to the total time. A full turn of the clock symbolizes the total time of a process. The precise elapsed time is displayed digitally below the clock. The clock runs

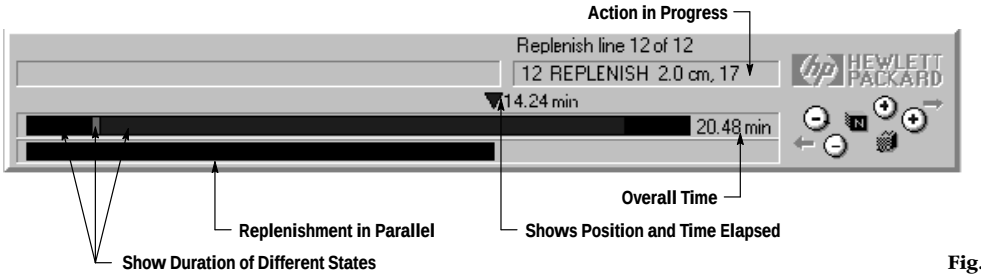


Fig. 7. Runbar

backward for actions taking a fixed period of time (e.g., flush for 3 minutes), thus showing the remaining time of the action. For procedures containing programmable actions (e.g., run), the clock runs forward, showing the time elapsed.

Measurement Unit

The cassette with the detection unit is the core of the user interface and the instrument (see Fig. 5). In the center of the user interface, everything in direct connection with the capillary, and thus the separation, is displayed. One can see whether the lamp is switched on or off, whether voltage or pressure is applied to the capillary and which vials are in contact with the capillary. The temperature inside the capillary cassette can be read. The pressure bottle symbolizes the pressure system. The black connecting lines show the existing connections to the capillary or the vials. They change to red when power is applied, to blue when compressed air is supplied, and to pink for vacuum. The pressure applied is displayed in the pressure bottle. Thus, it is quickly visible what is switched on at a given point in time.

Replenishment

The replenishment system display shows which vials are presently filled or emptied. For examples, see Fig. 6. Pressure on the lines is shown in blue, vacuum in pink.

Runbar

The runbar (Fig. 7)is designed to provide an overview of the entire analysis. With its help, the user can see how far the analysis has proceeded. An analog bar displays the relative durations of the individual parts of the analysis in different colors.

The runbar gives the user the following information:

- Which step is presently in process (colored parts)
- What fraction of the total time is consumed by each individual step
- Completion of the analysis (overall time)
- Time since the start of the analysis (triangular arrow)
- Position of the current step in the analysis
- Individual steps executed
- Whether parallel replenishment is active.

This helps the user get a clear overview of the state of the analysis. A short glance at the analog display suffices for a qualitative impression.

Detector and Electric State

The current values of the detector and the electrical unit are shown on the left side of the user interface. In the electrical parameter display (Fig. 8) voltage, current, and power are displayed and can be controlled. Whether voltage, current, or power control is enabled is indicated by a bold digital numerical value and a black bar (actual value) obscuring the corresponding green bar (reference value). The camera symbols show where data is being recorded. The display of the electrical values is located opposite the battery symbol.

The detector settings are shown in the user interface opposite the detector symbol (Fig. 9). Here, the settings for all channels can be read. Channels for which data is being recorded are indicated by camera symbols.

Online Plot Area

In the online plot area (Fig. 10) are two signal windows in which the signals from the detector or the electrical unit can be displayed. The displays show the values presently measured.

Upon the click of a button, either plot window can be enlarged so that more precise values can be read and the curve

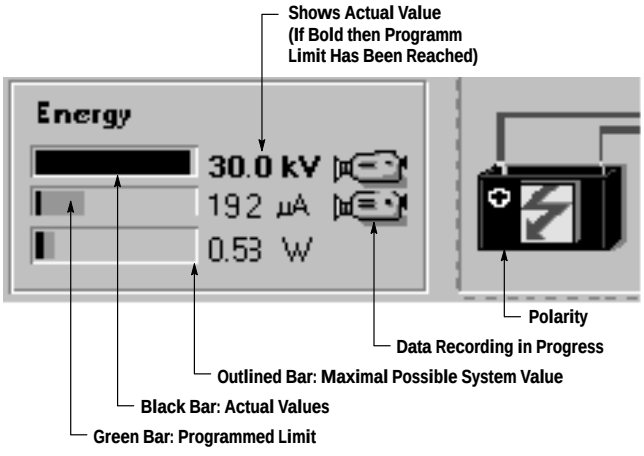


Fig. 8. Electrical parameter display.

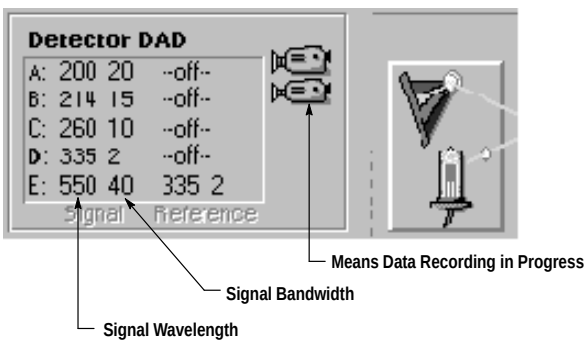


Fig. 9. Detector display.

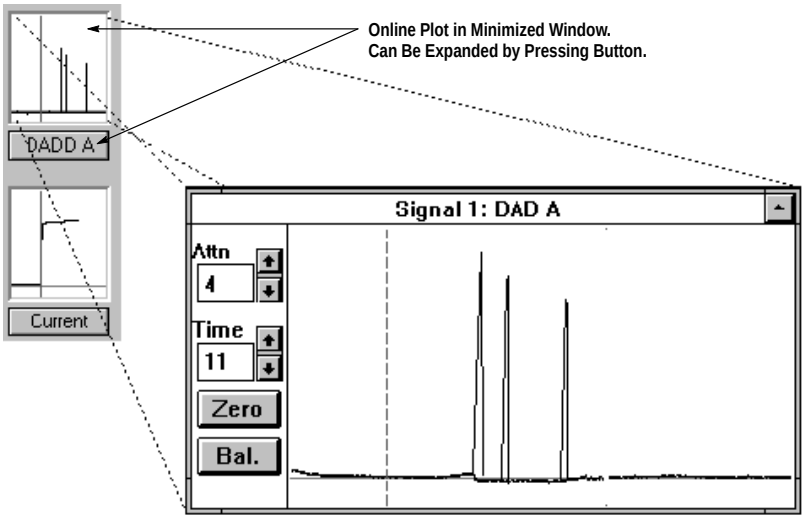


Fig. 10. Online plots and enlarged plot.

trace can be better followed. Clicking on the enlarged signal enables tracking of the signal curve and precise determination of individual values.

Simulation of a Method

Simulation is used for verification of a programmed method. The complete sequence of a method can be shown in the user interface. All actions are displayed as if the instrument were running. The simulation is software-based and does not require any hardware.

When the simulation is initiated, a control dialog appears which facilitates navigation through the simulation. Moving forward and backward in single steps is possible. With each step, the corresponding time at which the action is executed is displayed. The simulation only shows steps causing a change of state. A run button makes continuous stepping possible.

By simulation, the user can verify the complete sequence of a method before the actual run to make certain that the method is error-free. Especially in the case of methods with long run times, it makes sense to check in advance whether everything will go as planned or if program faults have slipped in. The user can also see in advance how long the analysis and the individual steps will take, and whether there are any conflicts.

Acknowledgments

We would like to thank Chris Wiesner and Craig Blackwood who helped us use their user interface library at the beginning of our project. Thanks, too, to the ChemStation teams whose code we reused.

Reproducibility Testing of the HP CE Instrument

The final chemical test developed for the HP CE instrument implicitly checks various instrument functions by determining the reproducibility of migration time and peak area measurements for well-defined chemical samples. The injection type was selected by testing four different types in a series of reproducibility tests. The final test can be used in production, at a customer site, or for teaching CE basics.

by Ulrike Jegle

Instrument tests are indispensable for quality assurance of highly integrated systems. Only tests of the whole system guarantee a quick setup of the instrument at the customer site and a minimum failure rate in daily use. Although the components of the HP CE instrument are tested separately for functionality, the integrated system has to be tested for successful coexecution of all built-in elements. Besides functionality tests, the procedure for the HP CE instrument includes measurements of detector properties (noise detection) and reliability determination within a reasonable test time.

For the final test of the HP CE instrument, a test procedure was implemented that exercises the instrument under real-life conditions. A chemical test procedure had to be developed for this. Besides testing the instrument functions, other important aspects had to be considered for this chemical test. The final test had to be fast and easy to execute and use commercially available and nontoxic chemicals. In addition, it was desirable that the test be usable as a customer-site test. Teaching aspects were also considered in choosing the test separation procedure, so that the test could be used to show the basics of the relatively new capillary electrophoresis technique. The advantages of dual use of the final test are cost reduction and, more important, the possibility of obtaining comparable results in final test and customer tests for validation reasons.

Test Description

For the chemical final test, bare fused silica capillaries are used to separate a set of three vitamins—thiamine (positively charged), nicotinamide (no charge), and nicotinic acid (negatively charged)—using a 20-mM sodium phosphate buffer of pH 7 as run buffer. The vitamins come in a lyophilized mixture that has to be dissolved in water before use.

To complete a high number of runs within a reasonable test time, short preconditioning and separation intervals are required. Short run times can be achieved by high electro-osmotic flow and high analysis temperature.

Table I
Conditions for Chemical Final Test

Capillary:	Bare fused silica Total length: 48.5 cm Effective length: 40 cm Inner diameter: 50 μm
Conditioning: (pretreatment before first use)	Flush: 10 min 1M NaOH Flush: 2 min 0.1M NaOH Flush: 3 min run buffer All steps carried out at 60°C
Run Buffer:	20 mM sodium phosphate pH 7
Solute:	1 mM each thiamine, nicotinamide, nicotinic acid dissolved in water
Preconditioning: (before every run)	Flush: 2 min 0.1M NaOH Flush: 3 min run buffer
Separation:	Voltage: 25 kV Polarity: positive Temperature: 40°C
Detection:	Wavelength: 215 nm Bandwidth: 16 nm Reference: off
Data Analysis:	Initial threshold: −2

To guarantee a high electroosmotic flow, the capillary is pretreated before its first use with 1M sodium hydroxide at 60°C. Treatment of the fused silica wall with 1M sodium hydroxide solution at an elevated temperature results in a maximum of charged silanol groups on the surface, which is the basic prerequisite for high electroosmotic flow. A second rinsing step consisting of a 2-minute flush with 0.1M sodium hydroxide and a 3-minute flush with phosphate buffer is added to be as consistent as possible with the preconditioning step used before every run. The capillary conditioning with 1M NaOH is carried out once before the first use of the capillary for the final test and is not repeated

Table II
Factors Influencing Migration Time Reproducibility

Instrument Parameters

Constant Cassette Temperature
- Peltier Element
- Fans

Constant Separation Voltage
- Power Supply
- Connections

Constant and Reproducible Liquid Level in Replenished Vials
- Pressure System
- Valves
- Tube Connections

Chemical and Physical Parameters

Electroosmotic Flow (Zeta Potential†)
- Fused Silica Material
  - Bulk
  - Surface
- Buffer
  - pH
  - Ionic Strength
  - Type
  - Viscosity (strongly temperature dependent)
- Equilibrium State
  - Preconditioning
  - Buffer Type

† Zeta potential is the potential difference between the capillary wall surface and the bulk solution.

during the test. The preconditioning is a flush step carried out before every single run. Here weaker conditions are used: 0.1M NaOH and run buffer.

In addition to creating a high electroosmotic flow, the run time of the vitamin separation is optimized by using a run temperature of 40°C even though this impairs the separation efficiency. Increasing the temperature results in a decrease of buffer viscosity of 2% per degree. This causes higher flow velocity of the run buffer. The preconditioning and separation conditions are listed in Table I.

It has been shown (see "Final Test and Results" below) that the test mixture of thiamine, nicotinamide, and nicotinic acid can be separated with high reproducibility of migration time and peak area under the conditions listed in Table I. Typical relative standard deviation values calculated for six runs were below 1% in migration time and below 3% in peak area.

Test Parameters

The metrics collected in the final test are the reproducibilities (relative standard deviations) of the measured migration times and peak areas of the well-defined vitamin samples. The instrument parameters implicitly tested by determining time reproducibility are the stability of the cassette temperature and the accuracy of the replenishment system. The entire injection and detection system is tested by the determination of area reproducibility. Of course, the basic instrument functions like tray and lift movements need to work correctly during the final test for any results to be obtained.

Time and area reproducibility depend not only on instrument parameters, but also strongly on chemical and physical influences. For testing instrument functions, the test tool, which is the vitamin analysis in this case, should have constant parameters, ideally with no failure possibilities. However, variations and failures of chemical systems cannot be excluded. Therefore, failures and their impacts have to be well-known to distinguish between chemical and instrumental failures.

Tables II and III show instrumental and chemical/physical parameters having direct influences on migration time and peak area.

Time Reproducibility

The most important parameters influencing the migration time reproducibility of separations carried out in bare fused silica capillaries are variations in the electroosmotic flow and the degree of equilibration of the silica surface with respect to the buffer.

The Smoluchowski equation describes the parameters determining the electroosmotic mobility:[1]

$$\nu_{EOF} = -\frac{\varepsilon_0 \varepsilon_r \zeta}{\eta} = \mu_{EOF} E,$$

where ν_{EOF} is the electroosmotic velocity, ε_0 is the permittivity of vacuum, ε_r is the permittivity of the electrolyte solution, ζ is the zeta potential, η is the viscosity of the electrolyte solution, μ_{EOF} is the electroosmotic mobility, and E is the electrical field strength. According to the Smoluchowski equation, the electroosmotic flow is influenced by the zeta potential, which is affected by the properties of the fused silica material. The electroosmotic flow is also affected by the composition and pH of the buffer and the preconditioning solution.

Table III
Factors Influencing Peak Area Reproducibility

Instrument Parameters

Injection System
- Pressure
- Valves

Detection System

Chemical and Physical Parameters

Expulsion of Sample
- Single-Step Injection
- Injection + Postinjection
- Voltage Ramp

Reproducibility of Injection Plug
- Injection Volume
- Zero Injection Effect

Time Reproducibility
- $\int t \cdot \text{absorbance}$

pH and Concentration of Conditioning and Preconditioning Solutions. Fused silica capillaries are drawn under highly dehydrating conditions of high temperatures (around 1600°C) and dry inert gas (argon). When a capillary is treated with aqueous conditioning solutions the first time before use, silanol groups are created, a gel layer at the silica/liquid interface is developed, and the electrical double layer is built up (in the case of untreated fused silica, this consists of negative charges on the capillary surface and positive charges in the liquid near the surface, as explained in the article on page 6).

In general, NaOH flushes at elevated temperatures are used for conditioning, resulting in a maximum surface density of charged silanol groups. NaOH is known to dissolve fused silica strongly and is therefore expected to have the best cleaning effect on the surface. On the other hand, the use of NaOH is a very severe method that is expected to leave a rough surface which could create electroosmotic flow variations and local eddies. This can decrease the separation efficiency.[2] The impact of NaOH is not entirely clear at this time and the subject is still under ongoing research.

Acidic solutions like phosphoric acid are also used for conditioning, especially when the subsequent separation is carried out in acidic buffer media.

Before every run, capillaries are normally preconditioned for the subsequent separation for cleaning and equilibration purposes. This step can consist just of one buffer flush but often includes several steps with different flush solutions. Preconditioning solutions that differ from the run buffer influence the status of the surface. For example, when NaOH flushes followed by run buffer flushes are used for the vitamin analysis, the resulting electroosmotic flow and surface condition are different than if only run buffer is used for preconditioning. In general, the electroosmotic flow is higher using NaOH flushes. In such cases, the surface is in a quasistable state and not in an equilibrium state.

Buffer pH, Concentration, and Type. The separation buffer affects the density of charged groups and the electrical double layer at the surface as well as the thickness and composition of the gel layer.

The type of buffer determines the equilibration behavior of the surface and the buffer. Phosphate buffers especially are known to have a slow equilibration behavior because phosphate migrates into and out of the bulk silica, continuously changing the zeta potential of the wall as long as migration takes place.[3,4] Using phosphate buffer at low pH after initial NaOH cleaning, surface equilibrium adjustment times over 20 days can be observed. For citrate or borate buffers, the migration mechanisms as described for phosphate are not known. Equilibration seems to occur faster with those buffers.

Fused Silica. The manufacturing of capillary material and the process of capillary production are expected to have an impact on the zeta potential of the surface. The type and magnitude of the influence of these parameters have not been clarified yet and are still under ongoing research.

Area Reproducibility

Besides the reproducibility of the injection pressure and voltage, which are instrument parameters, the type of injection procedure is one of the important factors that determine area reproducibility. The zero injection effect (explained later) has a similar influence on area reproducibility.

Type of Injection. The most common injection procedure is just to inject a sample plug of a certain length either by voltage or pressure. For better reproducibility, the sample introduction can be combined with an additional buffer plug injection (postinjection) with or without a voltage ramp to start the separation.

Theoretically, the following considerations apply. When a separation voltage is applied to a capillary after injection, a heat pulse is created. This results in the volume expansion of the buffer and sample in the capillary. Volume expansion causes liquid to be pushed out of the capillary. If the sample plug is located directly at the end of the capillary, some sample is expelled. This process is a cause of low reproducibility. Under the conditions in which the vitamin analysis is carried out, this effect has been calculated to have a minor impact.[5,6] Experiments (described later) have shown that the geometry of the capillary end is much more important. Even if the capillary has an accurately cut rectangular edge and the polyimide is burned off at the ends, the microscopic edge properties differ from capillary to capillary. In addition, physical interactions between the surface tension of the liquid and the properties of the inner and outer surfaces of the capillary contribute to the inconsistency of injected and expelled sample amounts. These deficiencies of the capillary ends can be overcome by the introduction of a buffer plug after the sample plug, as confirmed in the experiments.

Zero Injection Effect. All injection procedures are subject to the *zero injection effect*. Zero injection is caused by capillary forces when a capillary is immersed in liquid. Reproducibility is mainly affected by the interaction of the surface tension of the liquid and the capillary inner and outer surfaces, and by the inner diameter and length of the capillary.

For a capillary with an inner diameter of 50 μm and a total length of 48.5 cm, the amount of zero injection plug can be expected to be around 0.2 mm. This amount is added linearly to each injection plug. The influence of the zero injection effect decreases with increasing injection plug length, but increasing the injection plug length strongly decreases the efficiency of every peak. Efficiency and zero injection are opposite effects. A general rule says that the injection plug length should not exceed 1% to 2% of the total capillary length.[1]

Reproducibility Tests

To examine the exact influence of zero injection and injection type, a test suite was set up consisting of four different three-vitamin analyses. The conditions, except those for injection, were held constant and were the same as used in the final test procedure. The exact injection conditions are described in Table IV. Fig. 1 shows the uncorrected absolute areas of all three vitamin peaks (thiamine, nicotinamide, and

Table IV
Injection Conditions for Four Three-Vitamin Analyses

Method	Injection Time (min)	Injection Pressure (mbar)	Postinjection Time (min)	Postinjection Pressure (mbar)	Voltage Ramp at Separation	Sample Plug Length (mm)
1	4.6	40	4	40	no	2.9
2	2	40	no	no	2.4 s/20 kV	1.3
3	2	40	no	no	no	1.3
4	zero	zero	4	40	no	zero injection plug: 0.2

nicotinic acid). Fig. 2 shows the relative standard deviations of the peak areas shown in Fig. 1. The method numbers in Fig. 2 correspond to those in Table IV. Six repetitions of each method (injection type) were carried out for the relative standard deviation calculation. In addition, the test set of the four injection types was repeated four times to show the reproducibility of each.

As expected, the highest relative standard deviation was calculated for zero injection. Typical relative standard deviation values were around 13% to 20%. Comparing the test conditions, the best reproducibility, with relative standard deviation values smaller than 3% in all repetitions, was realized using the postinjection procedure (method 1). As expected, postinjection was the best means of overcoming the disadvantageous impact of the many different forces at the capillary ends. Using no postinjection but a linear voltage ramp of 20 kV in 2.4 seconds led to a failure range of around 5% relative standard deviation. As expected from theory,[5,6] the voltage ramp showed minor impact in that case.

Final Test and Results

The final test procedure was chosen on the basis of the above tests and considerations to minimize the effects of the factors that degrade migration time and peak area reproducibility.

At the beginning of the test procedure the capillary is flushed with 1M and 0.1M NaOH and run buffer (capillary conditioning), independent of whether the capillary is being used for the first time or has been stored in buffer for a longer period of time. Afterwards, 25 vitamin separations are carried out for each test. The first 18 runs are for equilibration of the capillary to an acceptable degree. The nineteenth run is used as a calibration run for the reproducibility report. From the following six runs, calculations for time and area reproducibility are done. Injection is carried out in a two-step procedure consisting of a sample injection by pressure followed by a postinjection of run buffer, also by pressure. From a statistical point of view the total number of test runs (25) gives a high enough probability that any instrument failure will be discovered.

Time Reproducibility. Why new capillaries have to be conditioned before first use can be directly inferred from the theoretical aspects described above. Cleaning and the creation of high electroosmotic flow are the main reasons. But the capillary conditioning with 1M NaOH also has to be carried out for capillaries stored in buffer. Theory leads to the expectation that storage in buffer over several days should result in an equilibrium state of the surface. However, if the capillary is used directly after storage, a bigger and longer electroosmotic flow drift during test separations is observed compared to the drift obtained after the conditioning method-with 1M NaOH. Probably, the preconditioning with 0.1M NaOH before every run destroys the equilibrium reached during the contact with buffer only. But elimination of the 0.1M NaOH flushes before every run leads to a worse peak shape

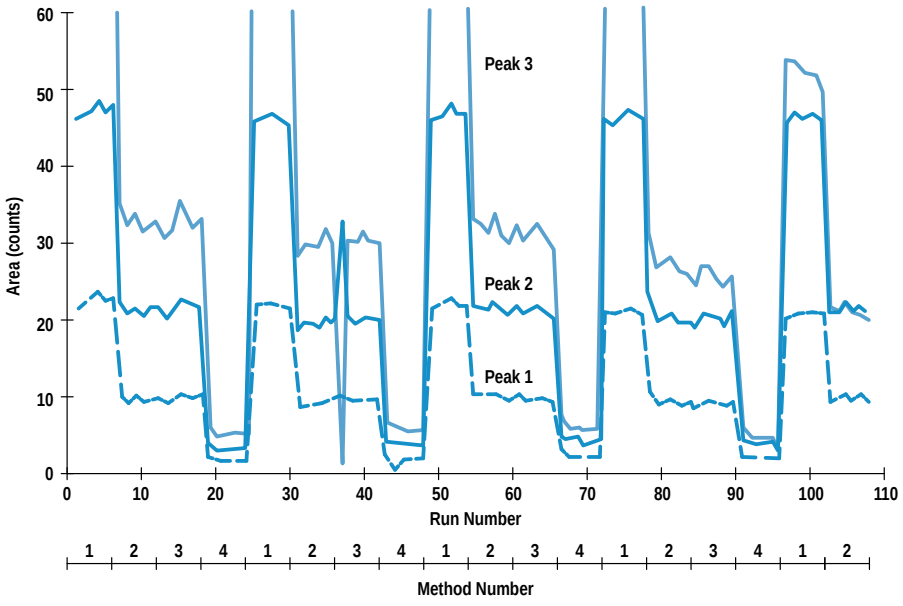


Fig. 1. Peak areas for thiamine (peak 1), nicotinamide (peak 2), and nicotinic acid (peak 3) using the four methods (injection types) described in Table IV. Six runs were done using each method and the entire sequence was repeated four times.

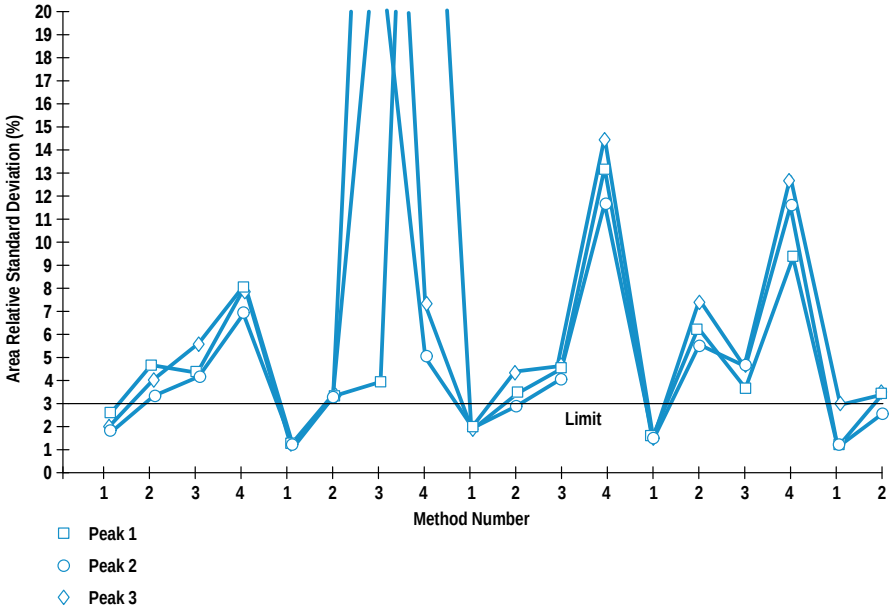


Fig. 2. Relative standard deviations of peak areas for thiamine (peak 1), nicotinamide (peak 2), and nicotinic acid (peak 3) using the four methods (injection types) described in Table IV. Six runs were used for the relative standard deviation calculation for each method and the entire sequence was repeated four times. The final test procedure for the HP CE instrument uses method 1.

for thiamine after a certain number of runs. In our tests, the best stability was reached by carrying out the capillary conditioning method once before the start of every test, followed by analysis procedures consisting of preconditioning with 0.1M NaOH and then run buffer, and finally separation.

Even when this procedure is used, the capillary needs 15 to 20 runs until stable enough conditions are established to ensure highly reproducible migration times. The procedure was verified in a series of five test sequences carried out using different capillaries. Figs. 3 to 5 show the migration times of the three vitamins thiamine, nicotinamide, and nicotinic acid plotted against the run number. Nicotinamide, the non-charged molecule, is only transported by electroosmotic flow. Therefore, this molecule can be used for the determination of electroosmotic flow. Figs. 3 to 5 clearly show variations within the first 15 to 20 runs which led to the deviations in migration times of all three test molecules. The impact of electroosmotic flow on negatively charged ions is much greater than on positively charged ions because negative ions migrate against the electroosmotic flow, so the influence of a fluctuating electroosmotic flow is much larger. This effect can be also deduced from Figs. 3 to 5.

Figs. 3 to 5 also show a comparison of the migration times of different sequences run with different capillaries. Differences between absolute values of electroosmotic flow for different capillaries are obvious. Electroosmotic flow variations up to 7% from capillary to capillary and up to 20% from batch to batch of bare fused silica could be observed. This is a well-known phenomenon described in the literature.[7,8]

Up to 1000 vitamin analysis runs can be carried out with a bare fused silica capillary treated as described above. In

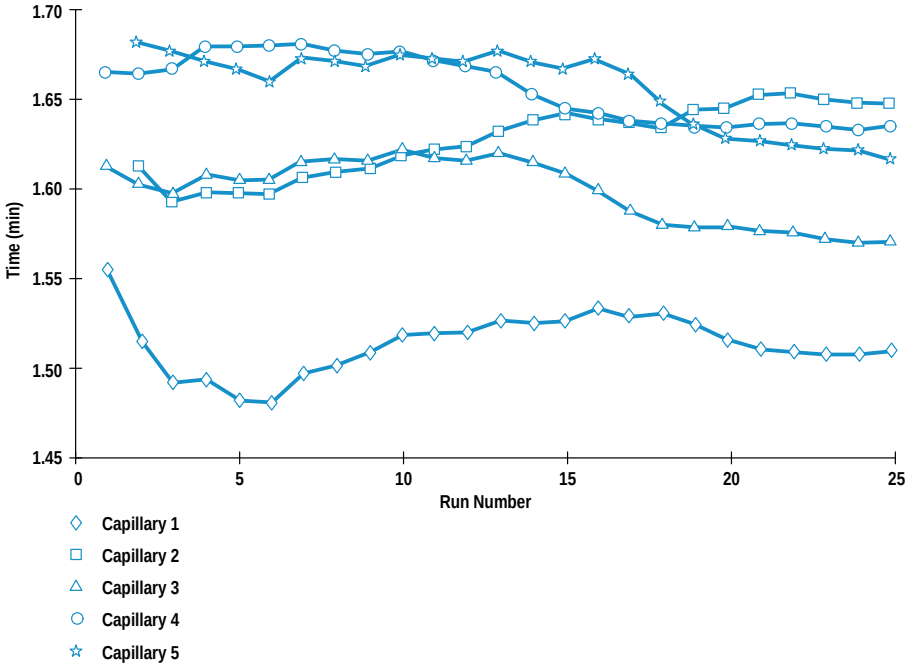


Fig. 3. Migration times for thiamine during the final test procedure for the HP CE instrument. The five plots are for five different capillaries.

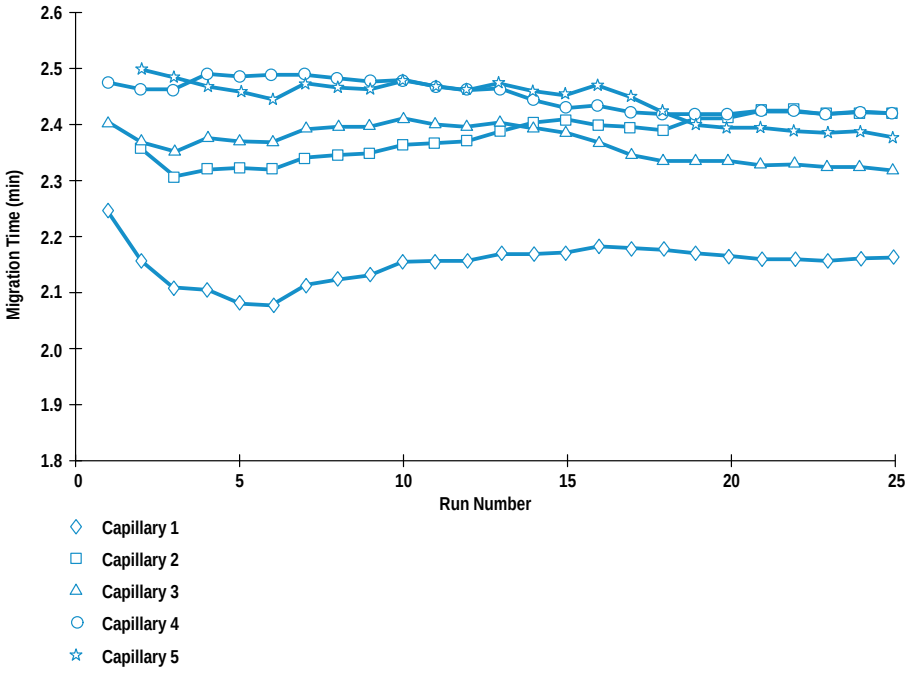


Fig. 4. Migration times for nicotinamide during the final test procedure for the HP CE instrument. The five plots are for five different capillaries.

general, a long-term drift to higher electroosmotic flow has been observed in long sequences.

Area Reproducibility. The injection procedure in the final test consists of two steps: the sample injection using the pressure mode is followed by a postinjection of run buffer by pressure as well. Both the sample plug and the buffer plug have a length of 3 mm. This plug length is a compromise between minimizing the zero injection effect and maximizing separation efficiency.

Another very important aspect for area reproducibility is the choice of the outlet vial during the injection procedure. The outlet home vial, which is the run buffer vial for the separation, has to be in position at the start of sample injection and must not be changed after sample injection. If the capillary outlet end is taken out of one vial and immersed into another, physical forces like surface tension of the aqueous liquid, inside and outer surface properties of the capillary, and pressure differences can cause the contents of the capillary to move. This can lead to either introduction of air or expulsion of liquid (sample) at the inlet side, the other end of the capillary. Like zero injection, this effect is hard to control and results in very high (failing) relative standard deviations. This problem is avoided by positioning the outlet home vial for all separations before injection.

Acknowledgments

The author wishes to thank Dieter Hass from quality assurance and Oswald Walczak and Ulrich Höger from production

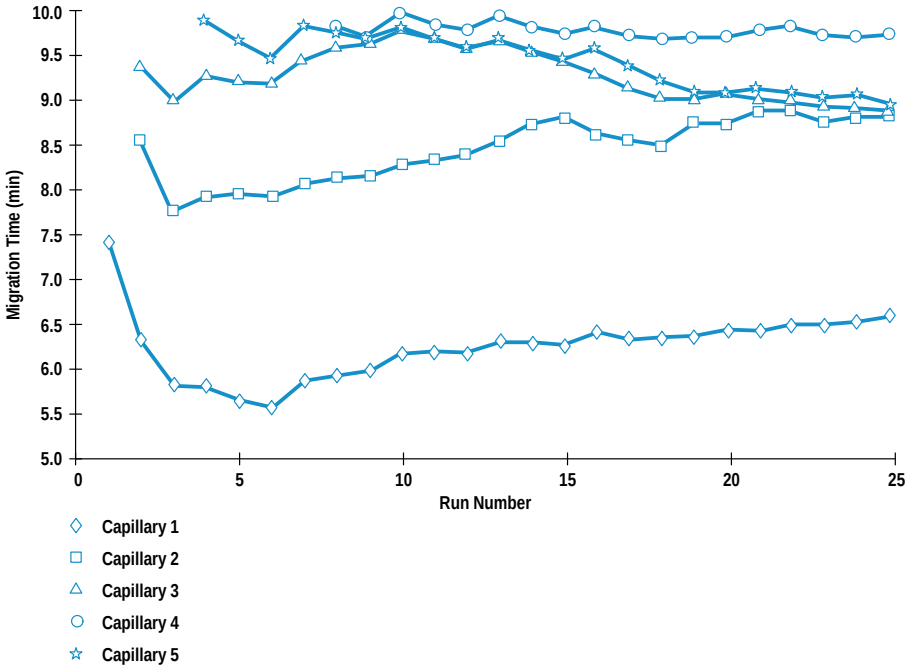


Fig. 5. Migration times for nicotinic acid during the final test procedure for the HP CE instrument. The five plots are for five different capillaries.

for their support in obtaining the measurements. Thanks also to Paul Wood for reading the manuscript of this article.

References

1. D.N. Heiger, *High–Performance Capillary Electrophoresis—An Introduction*, Hewlett–Packard Publication Number 12-5091-6199E.

2. S. Hjerten, "A new type of pH- and detergent-stable coating for elimination of electroendosmosis and adsorption in (capillary) electrophoresis," *Electrophoresis*, Vol. 14, 1993, p. 390.

3. R.M. McCormick, "Capillary Zone Electrophoretic Separation of Peptides and Proteins Using Low-pH Buffers in Modified Silica Capillaries," *Analytical Chemistry*, Vol. 60, 1988, p. 2322.

4. B.M. Mitsyuk, "Mechanism of the Reaction of Silica with Phosphoric Acid in Aqueous Solutions," *Russian Journal of Inorganic Chemistry*, Vol. 17, 1972, p. 471.

5. J.H. Knox and K.A. McCormack, "Temperature Effects in Capillary Electrophoresis, Part 1," *Chromatographia*, Vol. 38, 1994, p. 207.

6. J.H. Knox and K.A. McCormack, "Temperature Effects in Capillary Electrophoresis, Part 2," *ibid*, p. 215.

7. P. Coufal, et al, "The Magnitude and Reproducibility of the Electroosmotic Flow in Silica Capillary Tubes," *Journal of High–Resolution Chromatography*, Vol. 17, 1994, p. 325.

8. P. Coufal, "The Magnitude and Reproducibility of the Electroosmotic Flow in Silica Capillary Tubes," *International Conference of Riva del Gada*, 1993.

The Impact of Column Technology on Protein Analysis by Capillary Electrophoresis: Surface Coatings and Analytical Approaches for Assessment

To avoid unwanted interactions between proteins being analyzed and the surface of the fused silica CE capillary, the surface must be deactivated. Four approaches to surface deactivation for protein analysis are presented. A method for determining the extent of protein adsorption is discussed.

by Sally A. Swedberg and Monika Dittmann

In 1986, Henk Lauer and Doug McManigill of HP Laboratories published an article on the capillary electrophoresis of proteins on untreated fused silica.[1] In an attempt to solve the problem of untoward interactions of protein solutes with the high-energy fused silica surface, they took the approach of using organic additives in the electrolyte solution to deactivate the surface dynamically. They concluded that the approaches taken for dynamic deactivation were limiting for protein analysis in general for a variety of reasons. In conclusion, they wrote that "… the development of chemically bonded wall deactivators remains an important topic."

To give some insight into the impact of this study, it is important to understand what impact the CE project at HP was perceived to have at this time. Because of a number of publications by J.W. Jorgenson and others,[2] CE was thought to have the potential to revolutionize the way analysis of biopolymers would be done in the future. Therefore, protein analysis by CE was considered extremely significant. The conclusion of Lauer and McManigill that still more work on the capillary chemistry needed to be done to provide routine analysis across a range of protein samples drew new attention to research on coating chemistries for surface deactivation of fused silica capillaries.

This article describes research at HP Laboratories and the Waldbronn Analytical Division on coating chemistries effective for protein analysis by CE.

In Search of the Biocompatible Surface

Based on biophysical studies of the interactions of proteins at interfaces, it is known that one attribute of a surface that is necessary for reversible protein/surface interactions to occur is that the surface must be highly hydrated. Though this condition is necessary, it is not sufficient. Further, selection rules for surface attributes that promote reversible, low protein/surface interactions are not well-understood. The complexity of the heterogeneous polymer known as the protein makes rational approaches to such selection rules quite difficult. In addition, the Gibbs free-energy interaction term for protein/surface interactions is also

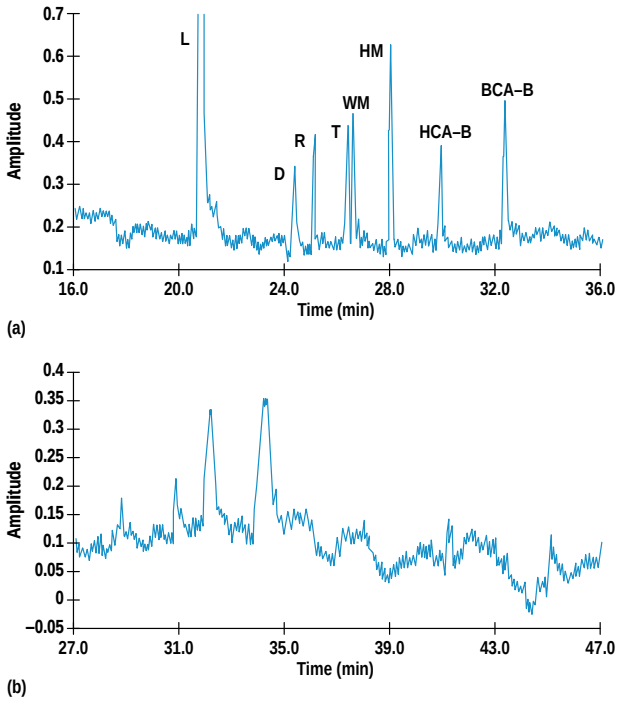


Fig. 1. Elution profile of seven protein markers and DMSO on (a) an arylpentafluoro-treated column and (b) an untreated fused silica column. Conditions: 200 mM phosphate, 100 mM KCL solution, pH 7, in 20-μm capillary tubing using an applied voltage of 250 V/cm. Detection: on-column UV at 219 nm. Length of column to detection: 100 cm. Peaks: L = hen egg white lysozyme, D = DMSO, R = bovine ribonuclease A, T = bovine pancreatic trypsinogen, WM = whale myoglobin, HM = horse myoglobin, HCA-B = human carbonic anhydrase B, BCA-B = bovine carbonic anhydrase B. (Reprinted from reference 3. ©1990 Academic Press. Reproduced with permission.)

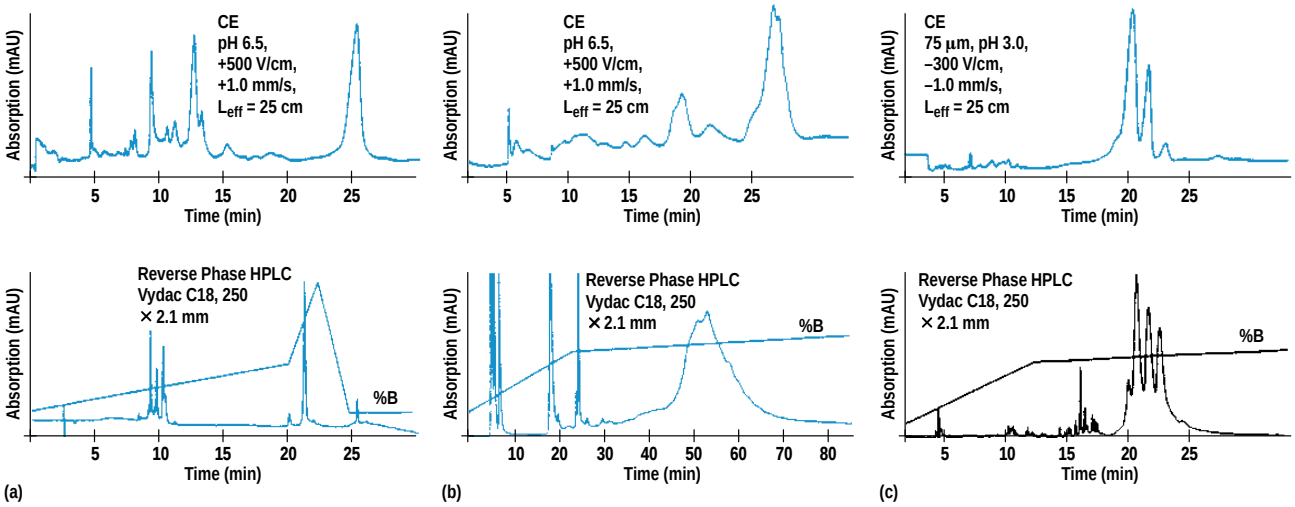


Fig. 2. Comparison of CE analysis on a bovine serum albumin modified capillary with HPLC analysis for (a) trypsin inhibitor, (b) ovalbumin, (c) α-chymotrypsinogen. With respect to the resolution of the number of species in the commercial protein preparations, the two methods appear to be macroscopically comparable, overall. In the case of trypsin inhibitor, CE separated more components than HPLC. For ovalbumin, HPLC gave more indication of the microheterogeneity in the major peak for this glycoprotein. For α-chymotrypsinogen, the leading shoulder peak of the major peak appeared to be more clearly resolved by LC. The total analysis time was considerably shorter for CE. Even in the case of α-chymotrypsinogen, CE does not require the reequilibration time required for reverse phase LC columns.

dependent on the solvent conditions used during analysis, not just on the chemistry of the surface alone.

Solving the problem of untoward protein/surface interactions requires both insights into biocompatible surfaces and a means for assessing the best solvent conditions in conjunction with a suitable surface. The next section gives an overview of some of the work done at HP Laboratories on surface coatings for analysis of proteins by CE. Then, work done at the Waldbronn Analytical Division to develop an analytical tool for assessing the extent of protein/surface interactions under various conditions is described. The specific example of a very promising surface, polyethylene glycol, is presented.

Surface Modified Capillaries

Arylpentafluoro Modified Capillaries. Fig. 1a shows the separation of seven protein standards on an arylpentafluoro-treated surface. This was one of the first surface modifications tried, and was based on attributes that were known from affinity and hydrophobic interaction chromatography (HIC) to be desirable for reducing protein/surface interactions.[3] Shown in Fig. 1b is the control, in which the same protein sample was run on a fresh fused silica surface. This separation, which is the result of untoward protein/surface interactions, is illustrative of the problems frequently encountered in the analysis of complex samples of proteins by CE on unmodified fused silica surfaces. As can be seen from the conditions, fairly high concentrations of phosphate buffer are required for this analysis. The second paragraph below discusses how the prolonged contact of phosphate with the silica surface produces an interesting surface modification that is also protein-compatible.

Bovine Serum Albumin (BSA) Modified Capillaries. Fig. 2 shows examples of protein samples analyzed by both CE and high-performance liquid chromatography (HPLC).[4] There are two interesting features of the BSA surface for the purpose of protein analysis. First, it is a highly hydrated surface, which in part explains the low sorptive nature of a variety of proteins to this surface, even at very low ionic strength of the buffer (10 mM citrate, in these examples). Second, since proteins are amphoteres, it displays a sigmodal electroosmotic flow profile as a function of pH (Fig. 3). It is therefore possible to tune both the magnitude and the direction of the electroosmotic flow as a function of the pH of the solution. The impact of this feature is demonstrated in Fig. 3.

Phosphosilicate Modified Capillaries. In 1988, McCormick demonstrated that with prolonged contact with phosphate buffers, a stable phosphosilicate surface is formed.[5] He demonstrated that successful protein separations could be done in fused silica capillaries after the phosphosilicate surface had been formed. His work was done at pH 2, which is an undesirable pH for many proteins, promoting denaturation and aggregation. Based on his studies, other groups have demonstrated that this approach can be used across a wide range of pH (Fig. 4).[6,7] Further, since these surfaces form readily, thereby obviating the need for specially manufactured capillaries, this is a useful method for protein analysis that can be easily integrated into the bioscience laboratory.

While physical means for assessing the extent of protein/surface interactions were used in conjunction with the separations discussed above,[8] a more robust approach using frontal analysis was developed at Waldbronn. This frontal analysis approach was used to assess still another promising surface, polyethylene glycol. That study is described in the following section.

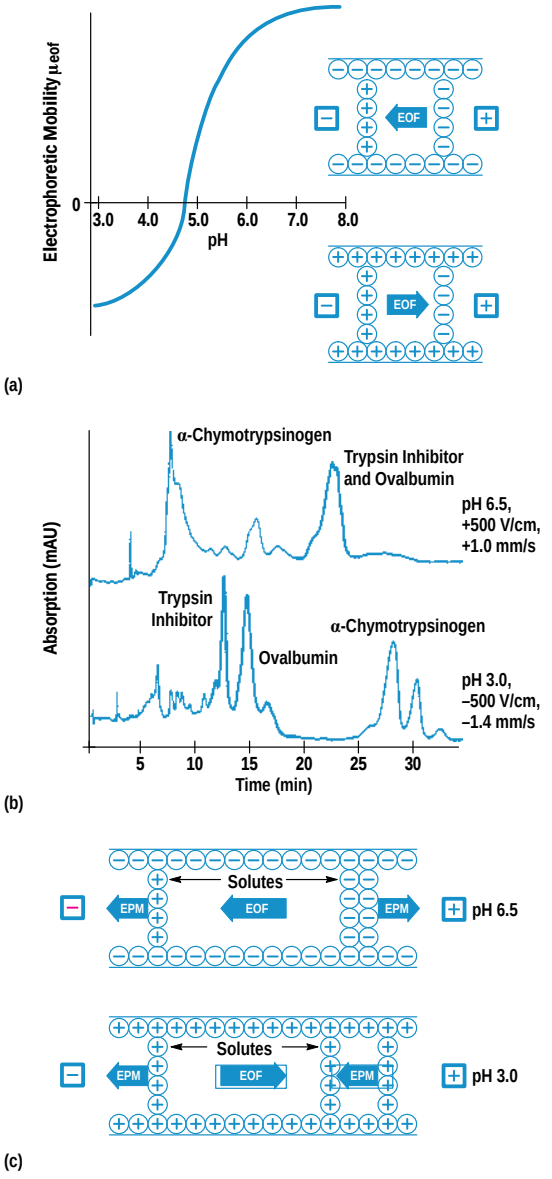


Fig. 3. Tuning selectivity of amphoteric phases. (a) Effect of pH on mobility. For amphoteric surfaces, the electroosmotic flow (EOF) changes predictably in both magnitude and direction as a function of pH. For the BSA surface the useful working range is between pH 3.0 and 8.0. (b) Control of resolution and elution order. For the same capillary, by changing the pH of the solution and the polarity of the voltage supply, it is possible to change the elution order and selectivity of the protein species. For these three proteins, pH 3.0 gives the best separation of the parent peaks.
(c) Since the net charge (both magnitude and sign) of a protein solute will also change as a function of pH, the solute electrophoretic mobility (EPM) changes with respect to the changing electroosmotic flow (EOF).

Polyethylene Glycol Surface Coatings

In 1989, the Waldbronn Analytical Division became involved in the development of biocompatible coatings for the CE separation of proteins. As previously mentioned, a surface that provides reversible protein/surface interactions must be

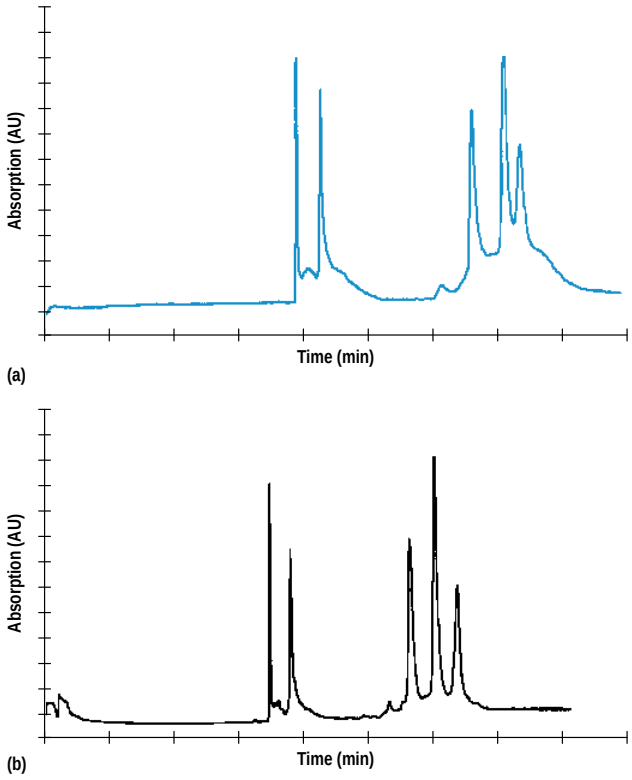


Fig. 4. Effects of sodium sulfate and ammonium sulfate on the separation of five protein standards. Buffer = 150 mM phosphate, pH 7.0, with (a) 200 mM sodium sulfate at 118 microamperes and (b) 200 mM ammonium sulfate at 138 microamperes. Conditions: 15°C, 400 V/cm, L = 49.5 cm, L_{eff} = 41 cm. Peaks (left to right): ribonuclease, myoglobin, B-lactoglobulin A, trypsin inhibitor, B-lactoglobulin B.

highly hydrated. One class of materials that fulfill this condition is the class of hydrophilic polymers known as polyethylene glycol polymers. Polyethylene glycol polymers as surface coatings are known to have little interaction with proteins.[9] Polyethylene glycol coated capillaries have long been available for gas chromatographic separations. This type of coating was tested at Waldbronn for feasibility in the separation of proteins.

Determination of Protein Adsorption by Frontal Analysis

A method that allows direct determination of the adsorption of a solute on a surface is frontal analysis. Frontal analysis is essentially a chromatographic method for measuring the amount of protein adsorbed and the strength of the protein/surface interactions. A solution of the protein of interest is pumped through the coated capillary and the breakthrough of the protein front is observed (Fig. 5).The amount of protein adsorbed at the surface under given conditions can be calculated from:

$$q(c) = c(V_f - V_d)/A_s,$$

where q(c) is the amount of protein adsorbed per unit area, c is the concentration of protein, V_f is the retention volume of the protein front (flow rate times retention time of the

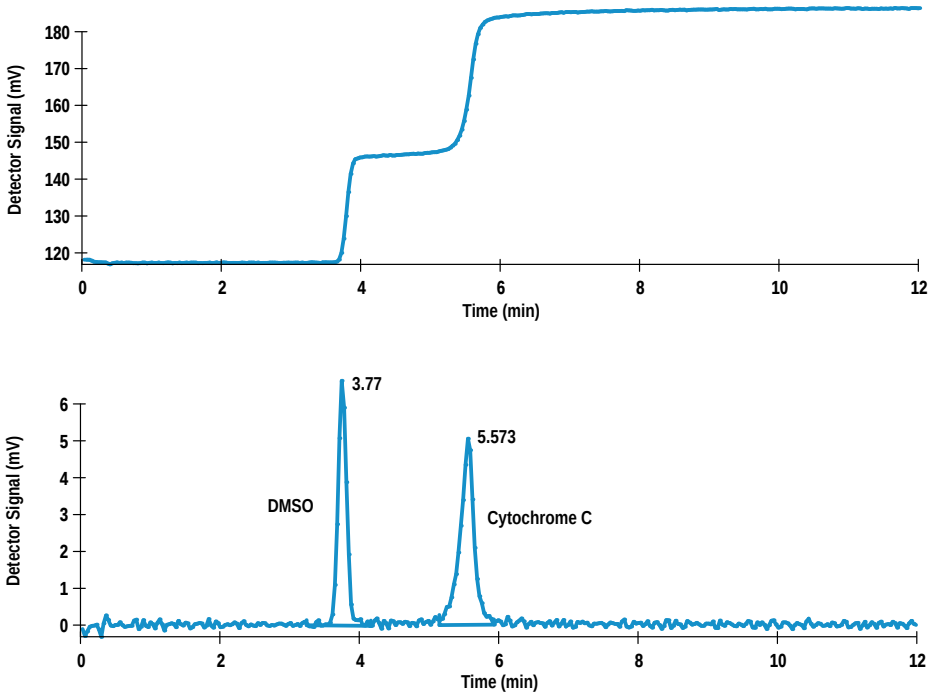


Fig. 5. Example of a frontal analysis curve of a mixture of DMSO (dead volume marker) and cytochrome C. The upper trace is the detector signal. The lower trace is the derivative of the detector signal.

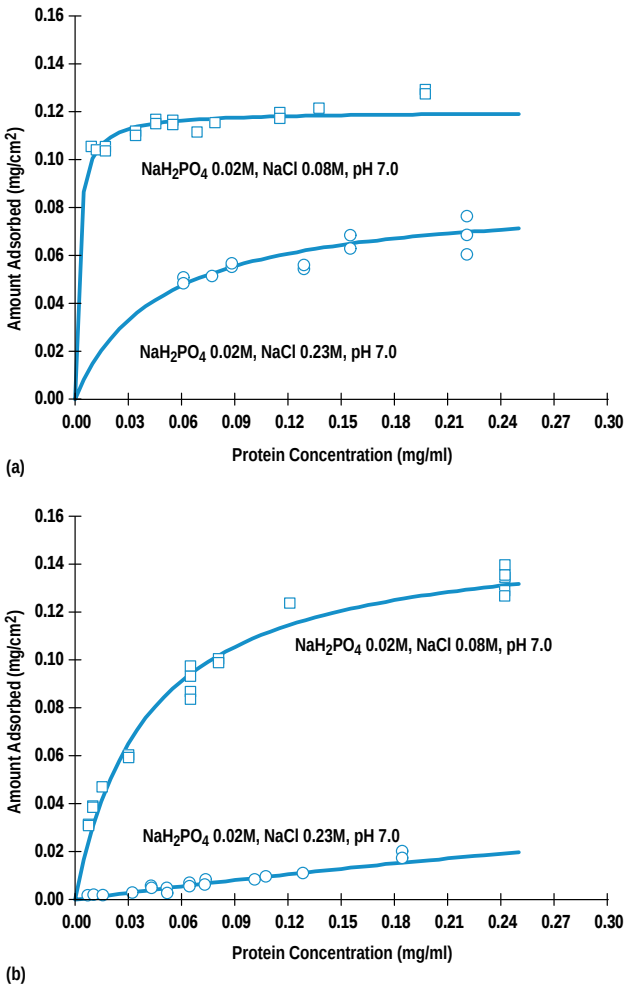


Fig. 6. Adsorption isotherms for (a) cytochrome C and (b) lysozyme on an uncoated fused silica capillary. Buffer: 0.02M NaH_2PO_4, pH 7.0, plus 0.08M NaCl and 0.23M NaCl.

protein front), V_d is the capillary dead volume (flow rate times retention time of the marker front), and A_s is the inside surface area of the capillary.

If q(c) is determined for a number of different protein concentrations, an adsorption isotherm is obtained. For many solutes, these adsorption isotherms are of the Langmuir type and can be fit to the equation:

$$q(c) = ac/(1 + bc),$$

where a is the slope of the isotherm at low concentration and b is an empirical parameter. The parameter a is related to the retention coefficient k′ by the equation $k' = a\beta$, where β is the phase ratio.

If adsorption isotherms of proteins are determined under different conditions (temperature, solvents, buffer concentrations, etc.) conclusions can be drawn as to the type and strength of protein/surface interactions. In this regard, proper separation conditions can be determined from frontal analysis experiments.

Adsorption isotherms of lysozyme and cytochrome C on an uncoated fused silica capillary at two different salt concentrations in the buffer (0.08M NaCl and 0.23M NaCl) are shown in Fig. 6. An increase of salt concentration reduces both the maximum amount of protein adsorbed at the surface and the initial slope of the isotherms at low protein concentrations. This indicates the presence of electrostatic protein/surface interactions on the uncoated fused silica surface that are suppressed by the addition of NaCl. Fig. 7 shows adsorption isotherms of lysozyme and cytochrome C on a polyethylene glycol coated capillary with no NaCl added to the buffer. On this surface, the amount of protein adsorbed is very small, indicating that this surface might be suited for protein separations. An addition of 0.03M NaCl completely suppressed protein adsorption on this surface.

Fig. 8 shows a separation of four proteins that cannot be eluted from an uncoated capillary unless a high pH or a very

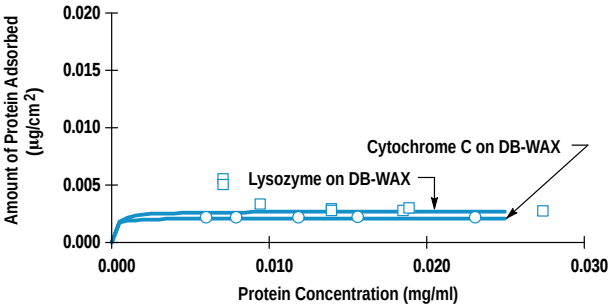


Fig. 7. Adsorption isotherms of lysozyme and cytochrome C on DB-WAX capillary at low protein concentrations. Conditions: 20 mM NaH_2PO_4, pH 7, 25°C.

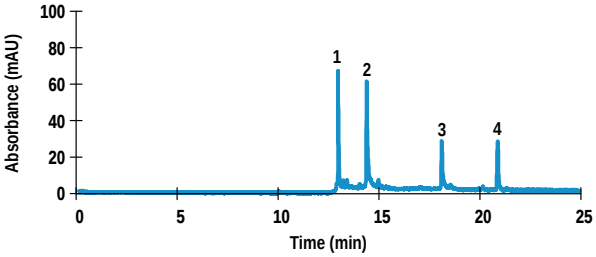


Fig. 8. Separation of (1) lysozyme, (2) cytochrome C, (3) ribonuclease A, and (4) chymotrypsinogen A on a polyethylene glycol coated capillary, L = 57.8 cm, L_{eff} = 51.3 cm, i.d. = 100 µm. Conditions: 20 mM phosphate buffer, pH 5.0, plus 30 mM NaCl, constant-voltage mode, 15 kV, 25°C, detection at 214 nm.

high salt concentration is used in the eluting buffer. On the polyethylene glycol coated capillary, a highly efficient separation can be obtained under moderate conditions.

Further studies concerning the stability of the coating and separation of real-life samples are under way.

Acknowledgments

The authors would like to thank Fred Strohmeier for his continued support and helpful discussions. Additional acknowledgments are expressed to Gerard Rozing and Doug McManigill for their helpful discussions and support on the frontal analysis experiments.

References

1. H.H. Lauer and D. McManigill, "Capillary zone electrophoresis of proteins in untreated fused silica," *Analytical Chemistry*, Vol. 58, 1986, p. 166.

2. J.W. Jorgenson and K.D. Lukacs, "Zone electrophoresis in open-tubular glass capillaries," *Analytical Chemistry*, Vol. 53, 1981, p. 1298.

3. S.A. Swedberg, "Characterization of protein behavior in high performance capillary electrophoresis using a novel capillary system," *Analytical Biochemistry*, Vol. 185, 1990, p. 51.

4. M. Herold and S.A. Swedberg, "Separation of contamination species in commercial protein preparations: a comparison of CE vs reversed phase HPLC," *Protein Society Meeting*, 1993.

5. R.M. McCormick, "Capillary zone electrophoresis separation of peptides and proteins using low pH buffers in modified silica capillaries," *Analytical Chemistry*, Vol. 60, 1988, p. 2322.

6. A.D. Tran, S. Park, P.J. Lisi, O.T. Huynh, R.R. Ryall, and P.A. Lane, "Separation of carbohydrate-mediated microheterogeneity of recombinant human erythropoietin by free solution capillary electrophoresis. Effects of pH, buffer type and organic additives," *Journal of Chromatography*, Vol. 542, 1991, p. 459.

7. S.A. Swedberg and C.A. Miller, *Protein analysis via phosphate-deactivated fused silica capillaries using salt additives*, Hewlett-Packard Application Note, publication number 12-5962-9965E, 1994.

8. D. McManigill and S.A. Swedberg, "Factors affecting plate height in high-performance zonal electrophoresis," in *Techniques in Protein Chemistry*, T. Hugli, ed., Academic Press, 1989, chapter 45.

9. A.B. Anderson, S.A. Darst, and C.R. Robertson, "Protein Absorption at Polymer Interfaces, " in *Proteins at Interfaces*, J.L. Brash and T.A. Horbett, eds., ACS Symposium Series 343, 1987, chapter 20.

A New High-Sensitivity Capillary Electrophoresis Detector Cell and Advanced Manufacturing Paradigm

By circumventing laminar flow while expanding the cross section of the analyte, this detector cell greatly increases both the sensitivity and the linearity of capillary electrophoresis. Manufacturing is made feasible by an advanced computer-controlled miniature lathe using machine vision.

by Gary B. Gordon, Richard P. Tella, and Henrique A.S. Martins

Capillary electrophoresis (CE) is a powerful and relatively new analytical technology used to separate and help identify complex aqueous mixtures of chemical and biological samples (see article, page 6). However, CE has often been characterized as a wonderful technology that is scaled too small. This criticism arises from the practical requirement to use very small capillary and sample sizes, which compromises sensitivity. Thus sensitivity is often the most important specification researchers use when comparing CE instruments.

Small Capillaries, Big Headaches

Capillaries for CE are usually 75 micrometers or less in inside diameter. With larger diameters, several adverse factors come into play. First, for a given electrical field, increased diameter means more internal heating in the electrolyte filling the capillary, causing an increased temperature rise. Second, the radial thermal conduction path, by which heat is dissipated, increases in length. As a result, the temperature at the center of the capillary rises even further. This causes a viscosity gradient, which allows molecules of the same species near the center to migrate faster, thus spreading the bands unnecessarily. As a practical compromise, a 50-micrometer inside diameter capillary offers a reasonable balance between sensitivity and resolution.

Sensitivity can be increased somewhat by injecting larger samples. But beyond a point, this practice is undesirable since it degrades resolution and broadens the peaks. For most users, the width of the injection is purposely kept narrow enough so that its contribution to peak broadening is perhaps 10% or less.

The shape of the peak at the detector is characterized by the mathematical convolution of three contributors: the rectangular injection plug, the Gaussian impulse response of the separation capillary, and the shape of the illuminated volume of the detector. Each contributor's spread can be characterized by its variance, a measure of the square of its width. The variance of the detected peak is the sum of the variances of the injection, the capillary, and the detector.

In practice, a typical injection might be one millimeter long and, when spread by diffusion during separation, might produce peaks around 4 mm wide. Such an injection contains only 2 nl of sample, 1000 times less than with liquid chromatography. If the sample contains only one ppm of the constituent of interest, the amount of this constituent that reaches the detector is truly minuscule.

Conventional CE Detectors

CE detectors usually measure UV absorbance. The problem becomes one of efficiently probing or illuminating the band of sample constituent carried in the transparent electrolyte. With 4-mm-wide peaks, a designer might choose to probe a volume of electrolyte 1 mm long to avoid excessive peak broadening.

A second important requirement of detectors is that they be able to quantify peaks accurately, that is, they must be linear.

Fig. 1. Inspiration from a 1903 handbook on glass blowing (from P.N. Hasluck, *Glass Working by Heat and by Abrasion*, Reprinted by Lindsay Publications, 1989; used with permission).

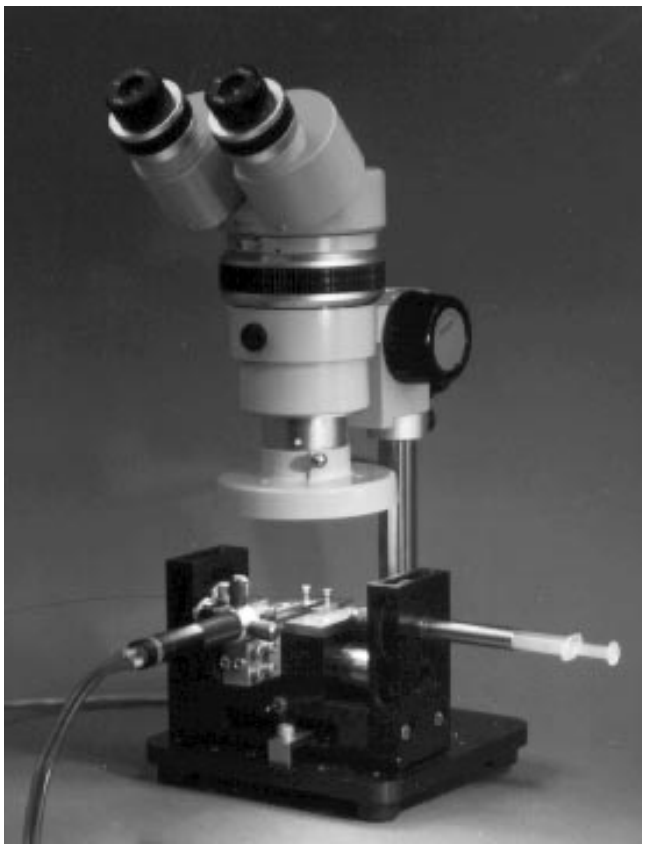


Fig. 2. Manual microglass lathe used for initial prototyping.

To accomplish this, as much of the illumination as possible should pass through the center of the capillary, and especially not travel as stray light around its edges. The intuitive reason for this requirement is that at high absorbencies less light makes it through the analyte, while stray light remains constant and becomes proportionally an increasingly significant contributor. Thus, further decreases in the already weak light passing through the analyte result in less-than-proportional changes in the total light received by the photodiode or spectrophotometer, that is, the detector becomes nonlinear.

Conventional CE detectors attempt to pass illumination radially through the capillary walls, focusing or aperturing the light as well as possible to pass the rays through the center. This configuration is fraught with problems. First, the absorbance path length is so short (the inside diameter of the capillary) that the signal is small and the sensitivity poor. Second, the laws of physical optics prevent most of the deuterium UV excitation lamp's energy from being focused into such a small region, so fewer photons reach the detector and the shot noise increases. Third, it is virtually impossible to prevent stray light from escaping around the capillary inside diameter, so the linearity is poor at high concentrations and quantification is inaccurate.

Lengthening the Path

To obtain a longer absorbance path length, it is tempting to imagine ways of probing the sample volume axially, such as with a Z-shaped cell, since it would better match the long aspect ratio of the bands. The benefit would be a path length of hundreds of micrometers, compared to one 50 micrometers long as in the foregoing example. The path length, however, cannot be the full 1-mm length of the probed volume as might be expected, since the light rays spend most of their time traversing the capillary walls as they zigzag off its periphery. One author briefly considered this option early on but rejected it because of poor linearity, high cost, and modest sensitivity gains.

Given the above difficulties, what other techniques might be used to increase the path length? Would radial expansion to form a bubble be feasible in a capillary not much larger than a human hair? If so, what might be the mixing consequences? These were the starting thoughts for the HP design, influenced by remembrances such as the glassblower sketch, Fig. 1.

To explore the nature of microglassblowing, first a CO_2 laser was tried, with unpleasant results (the capillary pitted and fractured). Next a small battery-powered glass lathe was constructed (Fig. 2.). Commercial pin vises were adapted to form headstock and tailstock clamps, which supported the capillary on both sides of what would become the detector cell. A small syringe, which rotated with the lathe, was used to apply an air pressure of several bar to one end of the capillary. The other end of the capillary was left open and allowed to spin freely. Fused silica softens at a very high temperature (1500°C), but a hobbyist miniature oxyacetylene torch worked very well for applying pinpoint heat. A stereo microscope facilitated watching the bubble growth progress. With this lathe, blowing the bubble cell proved to be relatively easy, despite some dire predictions. In fact, an important component of selling the project was having managers try it themselves.

The Bubble Cell: Laminar Flow, Plug Flow, and Serendipity

The core issue and more interesting science came to play when we began to consider what would be the nature of fluid flow in the cell. The first intuition was that the cell might represent a mixing volume, with either laminar or turbulent flow. With laminar flow, the velocity profile develops a parabolic cross section, with the innermost portion of

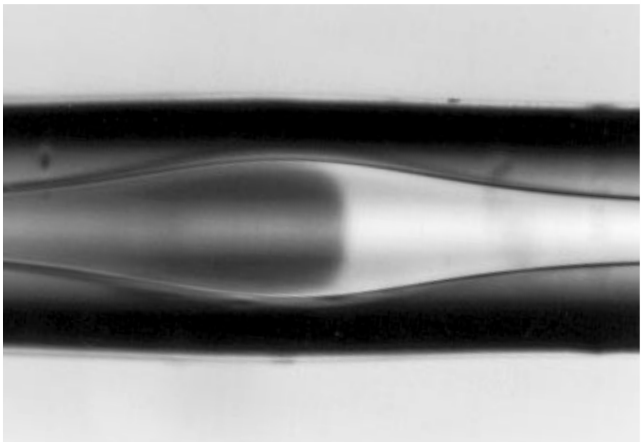


Fig. 3. Microphotograph of the plug flow front, confirming the absence of mixing.

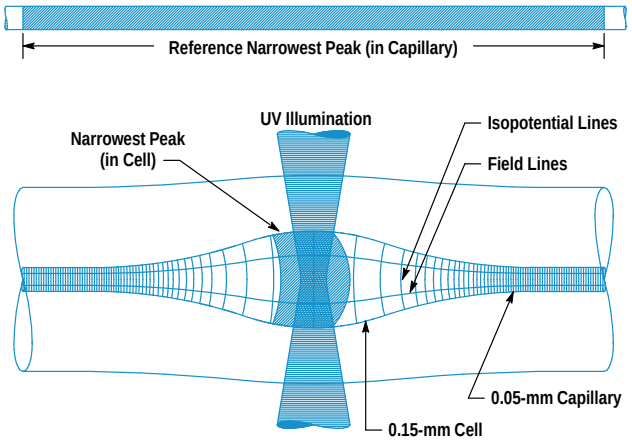


Fig. 4. Electrical field model of the cell.

the stream moving the fastest, and shear smoothly reducing the velocity at the walls to zero. Within this regime mixing and band broadening are considerable. A second and preferred form of flow in the cell is turbulent flow. With this form of flow, the cell can be thought of as a vessel that is repetitively partially refilled, mixed, and then partially re-emptied. With turbulent flow, $(1-1/e)$ or 63% of the volume is swept out with each complete refill. This is the regime used in liquid chromatography detectors, which are, of course, pressure-driven. It is not perfect, because the cell size must be much smaller than desired so as not to broaden the peaks excessively. As a consequence of the small cell size, the detector's sensitivity suffers.

As the project progressed, it became evident that a serendipitous flow balance in the cell might be conceivable. Unlike LC, the CE electrolyte is not driven through the cell by pressure, but rather propelled at its edges by the same electroosmotic forces that drag the electrolyte through the separation capillary. These forces might be reasoned to be less in the cell than in the capillary, just at the same time that the bulk flow requirements in the widening cell would require velocities that were decreasing.

Electroosmotic flow theory teaches that the forces that propel the electrolyte through the capillary act at the thin periphery of the fluid, across a region just micrometers thick, known to electrochemists as the double layer. These forces act to shear this double layer, producing a velocity near the edge that is proportional to force and therefore proportional to the axial electric field. Since the axial field is much lower in the cell because of the greater conductivity of its enlarged cross section, the velocity near the edge of the cell drops dramatically from that in the unexpanded portions of the capillary.

When the electrolyte enters the bubble cell, it encounters a gently increasing cross section. Since the fluid flow is incompressible, the average velocity in the cell drops in inverse proportion to the cross-sectional area. Now, if the velocity at the edge of the cell should also drop in proportion, there would be no tendency for mixing in the cell and the flow would be of the highly desirable plug nature.

This is, in fact, exactly what happens (see Figs. 3 and 4). Both the velocity at the edge of the cell, which drags the contents, and the contents of the cell slow down in exact proportion. This remarkable result means that unlike LC, the flow in the CE bubble cell is of plug nature, without any additional dispersion whatsoever. Thus cells proportionally larger by nearly an order of magnitude than for LC can be used, and much of the sensitivity lost because of the smaller CE apparatus is regained. It is as though the sample plug, long and skinny in the capillary, both slows and transfigures itself into a disc upon entering the cell, thus making possible much higher analysis sensitivity. Subsequently, upon exiting the cell, the plug retakes its original form and speeds along its way.

Fig. 3 shows a photomicrograph of the cell with the front of a separation band passing through it. Although the analyte's constituents continue to separate in the cell as they do in the separation column, the nature of the flow in the cell is dominated by the aforementioned electroosmotic flow.

The cell might, at first, appear to behave simply like a very large capillary and be vulnerable to radial heating, radial density gradients, and counterflow, which cause disastrous band broadening in large capillaries. This is not the case, in part because the the axial power density in the cell is a tiny fraction of the axial power density in the separation portion of the capillary. Further, any counterflow that might tend to develop is small because of the low field and the relatively short time the constituents spend in the cell.

In practice, a typical bubble cell might have a diameter three times that of the capillary within which it is formed. It is then said to have a bubble factor (BF) of three or 3×. In a 3× cell the osmotic velocity is one-ninth that in the separation region of the capillary. Peaks compress in length by a factor of nine, with a 4-millimeter-wide peak shrinking to only 440 micrometers.

How Sensitive?

The sensitivity gain of the bubble cell comes from two contributions: increased signal and less noise. Signals increase in direct proportion to the increase in the length of the absorbance path, which is equal to the bubble factor. We assume that the modern deuterium lamp is itself not a significant source of noise. Rather, the dominant noise derives from the randomness of the photons reaching the detector. Such "shot noise," as it is known, is proportional to the square root of the flux.

As an example of the potential sensitivity gain, consider a 2.5× cell compared to a straight 50-μm capillary. For maximum light flux and minimum stray light, assume that the round source of a deuterium lamp is demagnified and imaged directly onto the analyte. The signal increase is equal to the path length increase, which is equal to the bubble factor, 2.5. To calculate the decrease in noise, it is necessary to know how much the flux increases. Allowing a 10% margin at each side of the cell to minimize stray light effects on linearity, the beam diameter in this example could be increased from 40 μm to 100 μm, or a factor of 2.5. Since the flux increases by the square of the diameter, and the noise drops by the square root of the flux, the noise simply drops by the same bubble factor of 2.5. The signal-to-noise improvement is equal to the product of the path length increase and the noise reduction factor. Since each is

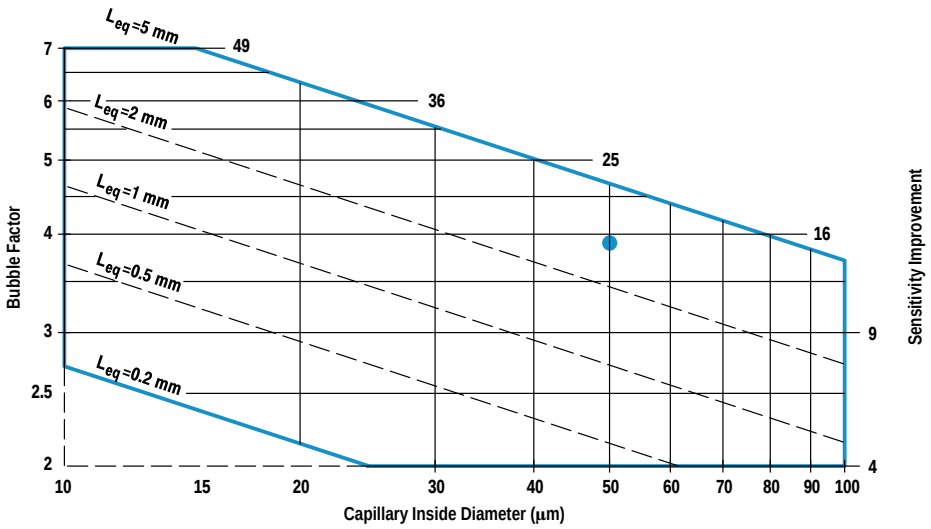


Fig. 5. Nomograph characterizing a bubble cell detector illuminated with a round illumination cone of optimized size and constant intensity. For a given capillary size and equivalent sampled length L_{eq}, an optimum bubble factor is indicated and the sensitivity improvement is predicted. The dot at a capillary diameter of 50 µm and L_{eq} = 3 mm indicates parameters commonly used in the CE literature to compare high-sensitivity detectors.

itself equal to the bubble factor, the signal-to-noise increase, or sensitivity gain, is simply the bubble factor squared, or 6.25× in this example.

Fig. 5 is a nomograph showing to the first order how several parameters are affected when a bubble is added. It assumes imaged illumination as above, but neglects the 10% margins. The dotted lines indicate the size of the sampled volume, as measured by an equivalent length L_{eq} of an analyte plug in the unexpanded capillary. For the above 2.5× cell example, the equivalent sampled length is seen to be 0.8 mm, a length unlikely to contribute noticeable peak broadening. Equivalent sampled lengths larger than roughly a third of the length of the separated peak (caused by too large a bubble) are less desirable because further sensitivity gains are less than proportional and peak broadening will occur. This point is approached with a bubble factor of 3 with 50-µm capillaries and bubble factors of 4 to 6 with small capillaries, using the illumination described.

In practice, demagnifying the source directly onto the cell requires careful and individual alignment. Alternatively, overfilled apertures at the cell can be employed. Since the apertures of the unexpanded and bubble cell capillaries could conceivably be made equal in area, the standard cell and the bubble cell would seem to have the same noise component. Thus the bubble cell would have only the signal boosting advantage, yielding a sensitivity gain identical to the bubble factor. However, common source lamps are not spatially broad enough to image onto and illuminate the full length of the analyte plug in the capillary, but rather only a few tenths of a millimeter of it. Therefore, in practice, bubble cells even in this case offer some noise reduction as well as increased signal. The aperturing practice in general extracts a slight overall noise penalty by limiting the total flux reaching the detector to only those nearly parallel rays that actually traverse the cell.

Reference 1 compares numerous other sensitivity-enhancing paradigms, and as of 1993 cites 42 other references in this burgeoning field. Many of the high-sensitivity detectors it cites are compared at sample volumes of 50-µm diameter and 3-mm length, a point marked on the nomograph with a dot. A bubble cell with comparable volume is seen to have a bubble factor of approximately four and yield a sensitivity enhancement of 15×. Further improvements are possible with cells oval or hourglass-shaped in cross section.

Bubble cells are amenable to use with rectangular capillaries,[2,3,4] fluorescence detection, laser excitation, and spectrographic detection. Reference 5 explores the flow characteristics of bubble cells.

The bubble cell has been characterized here in one of its best configurations, whereas in practice the realized gains depend on the individual optical configuration. For example, the HP G1600A CE instrument adds several important capabilities: spectrographic detection, interchangeability between bubble and conventional detector cells, and apertured self-aligning drop-in cassettes (see article, page 20). HP conservatively represents the cell sensitivity enhancement as being equal to the bubble factor (for example, a 3× bubble increases the sensitivity threefold). Currently HP offers 25-µm capillaries with 5× cells, 50-µm capillaries with 3× cells, and 75-µm capillaries with 3× cells.

The bubble cell affords an easy and economical method to increase the sensitivity of CE. It also increases the linearity and dynamic range of CE, bringing it up to the quantification accuracy enjoyed by liquid chromatography users. Marginal sensitivity and poor linearity are no longer barriers to adopting CE as a powerful new analysis paradigm.

Decision to Automate

HP's original plan was to employ skilled operators to fabricate bubble cell capillaries on manually operated glass lathes of the kind described above. However, the control of dimensional tolerances for diverse bubble sizes, blown in capillaries of different diameters, was recognized as a substantial challenge, especially as production demand increased.

Fortunately, the original work took place in HP Laboratories, HP's central research facility. The climate of HP Laboratories encourages crossfunctional awareness of the diverse technical programs. As a consequence, the prospects of automating the bubble-blowing operation stimulated the interest of the manufacturing research assembly technology team (ATT). The ATT is chartered to explore and create flexible state-of-the-art assembly technology to solve critical, technically challenging, manufacturing problems.

An extended partnership was quickly formed, involving the ATT, the Analytical/Medical Laboratory, which pioneered the bubble cell detector, and HP's Waldbronn Analytical Division, which had product design and manufacturing responsibility.

Drawing heavily on previously created technology, the ATT carried the project from inception to manufacturing production in six months, with the investment of one engineer-year of effort. The remainder of this section describes BubbleWorld, the technical solution, and in particular the crucial role played by computer vision.

As a starting point, given the proven capability of the manual lathe, the ATT determined to leverage its basic configuration. The ATT has many years of application experience employing computer vision as a prime sensory modality for feedback control of robotic assembly systems. It was, therefore, natural to choose computer vision as the vehicle for real-time process monitoring and control and for postfabrication inspection and dimensional characterization.

Specifications and Goals

Specifications and goals were established as follows:

- Capillary initial inside diameter: 25 to 75 μm
- Customizable bubbles: Variable bubble sizes and shapes
  Programmable axial stretch
- Polyimide coating stripping: Flame removal
- Bubble factor: 1× to 5×
- Postinflation axial stretch: 0 to 100 μm
- Reproducibility: ±5 μm on the inflated inside diameter
- Throughput: 2 units/minute
- Hands-off operation, except for loading and unloading
- Automatic quantitative process characterization and logging.

BubbleWorld Hardware

A close-up view of the business section of the microlathe in operation is shown on the cover of this issue, while Fig. 6 illustrates an overall engineering drawing. All of the lathe operations with the exception of capillary loading are under autonomous computer control.

The lathe consists of a fixed headstock and an axially positionable tailstock. The tailstock rides on a preloaded cross-roller assembly, and is positioned by one of four dc servo motors through a cam-link arrangement. This configuration facilitates adjustment to achieve a variety of positional resolutions and ranges, in addition to providing the requisite time response. Similar mechanisms are used to position the radial and axial stages for the torch. A fourth motor drives the spindle.

Identical head and tail spindles are driven by geared belts off a common jack shaft coupled directly to the spindle motor. The capillary is held concentrically in the spindles by two identical collets. Each collet has four fingers formed integrally with the spindle shaft. Collet design was complicated

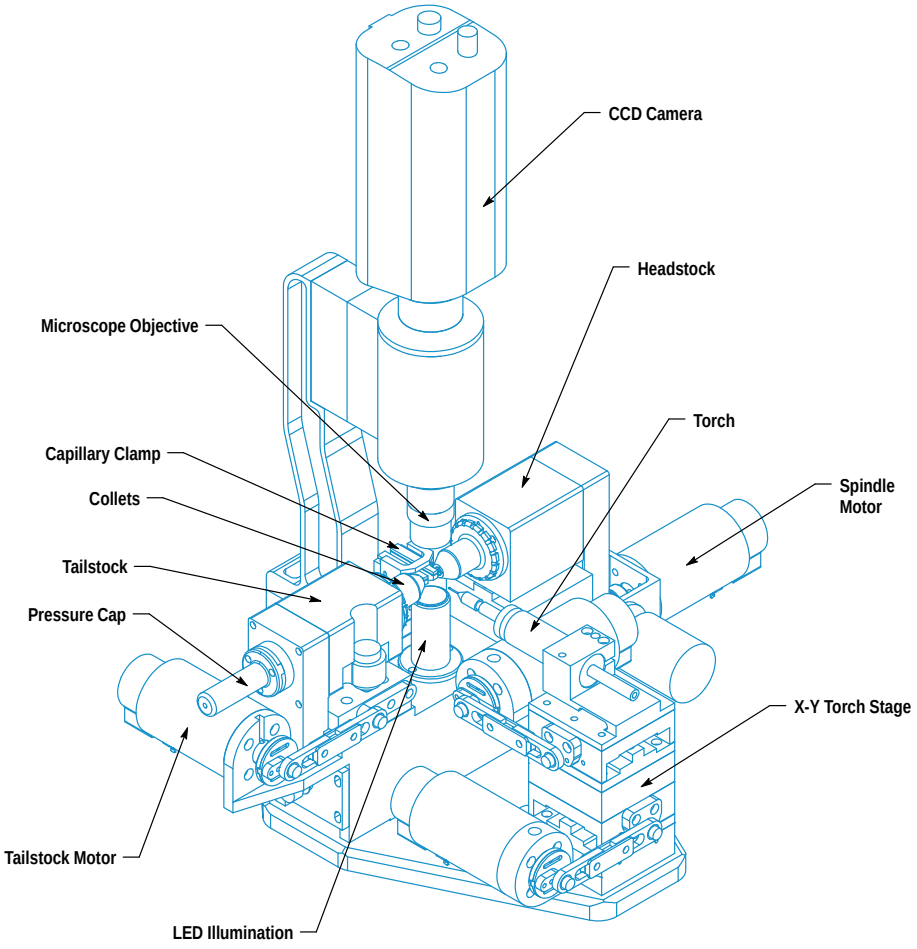


Fig. 6. Schematic drawing of the tabletop-size custom microglass lathe, nicknamed BubbleWorld, which incorporates four servoed axes and machine vision.

by the requirement for clearance to accommodate a stop-bead premounted on the capillary to reference the axial position of the bubble when installing the capillary in the CE. The stop-bead is located close to the bubble, just 4 mm from the collet finger tips, posing a headstock design challenge.

Each collet is actuated to capture the capillary by an outer cap that screws onto the spindle, imparting a radial closing action. A set of locking cylinders set into the stocks actuates to engage the outer cap to inhibit rotation. Proper sequencing of the lock cylinder actuators relative to spindle rotation enables automatic capture and release of the capillary.

Capillaries are inserted through the headstock into the tailstock. Insertion is arrested by the premounted reference stop-bead coming to rest against a registration lip in the headstock collet. The capillary is pressurized through a pressure cap in the headstock, 15 centimeters from where the bubble is to be blown. Since the capillary is one meter long and its flow impedance is large compared to that of the pressure delivery system, the opposite end of the capillary is left open to the atmosphere.

Coaxial alignment of the headstock and tailstock axes is critical to the fabrication of straight capillaries. Alignment is augmented by a small V-grooved finger configuration mounted between the collets. The fingers provide added support to the capillary during bubble blowing.

A miniature hydrogen torch is used, first to burn the polyimide coating before the bubble is blown, and then to heat the capillary during the blowing operation. The torch is mounted on a two-axis stage driven by servo motors and linkage arrangements similar to that used for the tailstock. Because the torch temperature is fixed during operation, the rate of thermal transfer to the capillary is controlled by a programmable radial positioning of the torch.

Proper distribution of heat along the capillary is achieved through programmable velocity control of the axial stage of the torch. This is generally achieved by a dithering motion of controllable amplitude and frequency. The vertical height of the torch is adjustable, but fixed.

A CCD camera fitted with a microscope objective views the capillary from above. High-contrast images of the inner diameter of the capillary are critical for measuring and tracking the bubble growth during blowing. A single diffused LED positioned behind the capillary provides high-contrast images of the inner and outer diameters. An infrared filter suppresses luminous glow from the hydrogen flame and heated capillary.

The workcell controller is an HP Series 9000 Model 720 computer. It runs the HP-UX* operating system and the X11 Window System, and performs all of the data processing, including real-time processing of the bubble images. The computer is equipped with a RasterOps VideoLive EISA-based I/O card used as an image frame grabber for the Cohu CCD camera, and to provide live image observation of the bubble-blowing operation. All motion control is achieved by a Galil four-axis motion controller commanded by the workstation through a serial line. All system sensors and actuators are accessed via the Galil controller as well.

Operation

Bubbles of customizable shape and size can be blown. The process development engineer can create and store numerous bubble-blowing recipes, each containing approximately 30 programmable parameters. Parameters define bubble profile, torch dithering, spin rates, image parameters, and the like. The lathe operator need only select a recipe.

Capillary loading is semiautomatic. The operator inserts the capillary into the open end of the headstock spindle (right side of Fig. 6) until the stop-bead is arrested by the registration lip inside the collet. The operator gently holds the capillary against the lip, while the machine locks the headstock cap and torques the spindle to tighten the collet. The tailstock collet is then similarly tightened. Once both collets are clamped, the supporting fingers are actuated, the synchronized collets are set into rotation, and the capillary is pressurized.

The capillaries are manufactured with a thin layer of protective polyimide. This coating is removed from the bubble section before blowing. This is accomplished by an axially dithered burn-off torching of the capillary. Following burn-off, the capillary is imaged to find the inside diameter. (Image processing is explained below.) The torch again approaches the pressurized capillary to heat it, and expansion commences. During expansion, the bubble image is tracked at approximately 10 frames per second. The output of the tracking operation is processed to support feedback control of the inflation.

If the torch were held at a fixed position during blowing, wall thinning at constant gas pressure would soon precipitate uncontrolled bubble growth. This would stress the ability of the system to control the final bubble size. Instead, in BubbleWorld, when the bubble achieves 80 percent of the goal, bubble growth is slowed by partially withdrawing the torch, allowing a controlled approach to the final diameter.

As an optional final forming option to allow additional control over the shape, the bubble can be axially stretched under program control by displacing the tailstock. Following bubble fabrication, its shape is thoroughly and accurately characterized. Lastly the capillary is removed by reversing the loading sequence. The entire cycle takes about 20 seconds.

Graphical User Interface

Control of the workcell is achieved via a graphical user interface (GUI) built upon a generic GUI developed at HP Laboratories for robotic workcell control.[6] The BubbleWorld GUI consists of a set of X11 windows, each dedicated to a particular task. The operator, the manufacturing engineer, and the maintenance engineer each access different GUIs. A representative set of GUI windows is shown in Fig. 7.

The top-level operator's interface consists of three windows: vision, statistics, and control. The vision window allows the operator to monitor the blowing operation, and includes graphical overlays representing vision system tracking of bubble growth. The statistics window reports bubble-blowing

performance history. The control window presents the operator with a list of predefined process recipes from which to choose.

At a second, password-protected, and more powerful level of control, production engineers can interact with the system to develop processes. This engineer's interface has several windows that allow control of the hardware. It includes a joystick window which allows independent control of each of the axes of the lathe and the torch XY stage. The engineer can control position, speed, and acceleration and the execution of bubble-blowing subtasks such as resetting the workcell, chucking and unchucking the capillary, and burning the polyimide coating. An output window allows the turning on and off of digital output bits and the setting of analog output channel values. An input window portrays the states of several digital input bits and analog channels. A communications window shows the commands sent to the Galil controller and can be used to send debug commands. A database window allows the engineer to store all pertinent parameters for a given recipe, including hardware parameters, and software parameters controlling the vision algorithms that track and measure bubble growth. A vision window accesses vision tracking and measuring parameters. Collectively, these various windows allow easy and intuitive interaction with the workcell, along with appropriate levels of security.

Computer Vision

As previously noted, computer imaging is used for feedback control of the bubble-blowing process and for postprocessing characterization of each bubble. The bubble image is fed from the microscope camera to an image frame grabber board which buffers the image for subsequent processing by the workstation. All image processing is performed by image processing software running on the workstation. We now describe the three vision operations FIND, TRACK, and MEASURE.

The FIND Operation. The purpose of FIND is to construct a matched filter to be used in the subsequent TRACK operation which monitors the evolving profile of the inner wall of the capillary as the bubble is blown. Fig. 8a shows an image of the backlit capillary after the polyimide has been burned away.

The first step in the FIND operation is to high-pass filter the image of the burned-off capillary. This step highlights cross-sectional dark-to-light transitions, as illustrated in Fig. 8b. Local maxima are referenced to a database of edge-pattern sequencing and diametral information. The result is the identification of the capillary walls in the image, as indicated in Fig. 8c.

From the high-pass filter data, a wall-tracking-filter matched to the dimensions of the particular capillary is automatically constructed. The FIND operation is designed to tolerate a degree of spatial noise such as that evident in Fig. 8, as well as noise caused by an exceptionally dusty imaging array.

The TRACK Operation. The TRACK operation tracks the growth of the inner wall of the capillary. The output of this operation is fed back to control the precision blowing of customized bubbles. To support real-time feedback control of bubble-blowing, it is desirable to process bubble images at a rate of at least 10 frames per second. The general-purpose computer is not able to apply TRACK over the entire image at this rate, but neither is this necessary; it suffices to process only a half dozen transverse sections.

The TRACK sequence consists of the following steps. First, the matched filter derived in the FIND operation is used to

Fig. 7. A custom graphical user interface (GUI) was designed to allow the user to develop, execute, and monitor the manufacturing process easily. There are windows for setting up parameters, collecting statistics, and displaying live video of the forming bubble.

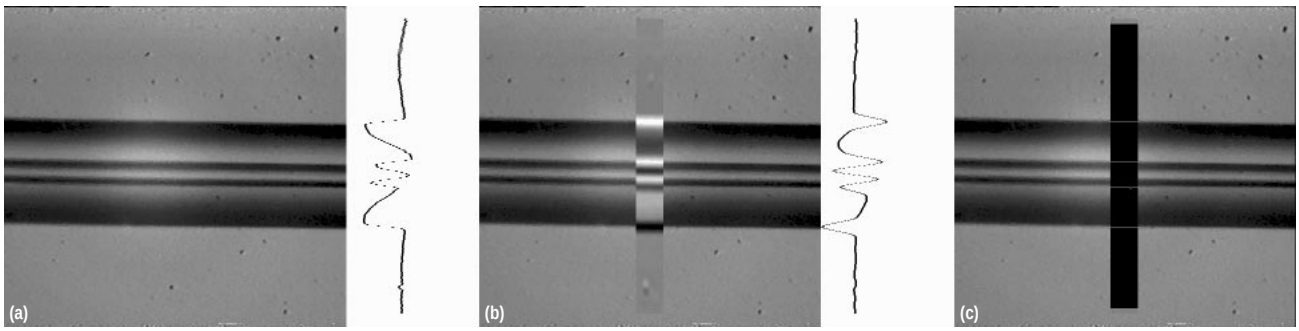


Fig. 8. Machine vision and tracking software are key to automating bubble cell manufacture. (a) The FIND step builds a matched filter to locate the inside diameter of a fresh capillary. (b) A high-pass filter processes a cross section of the capillary image to enhance its edges. (c) Peak detection is used to pinpoint the exact inside diameter.

locate the center of the capillary on both sides of the region where the bubble is to be blown. The central axis of the capillary is then defined by a line connecting these two centerpoints.

The image is then analyzed at several equally spaced transverse segments along the axis of the capillary. At each of these segments, edge and peak detection operations similar to those applied in the FIND operation are performed. The peaks associated with the first dark-to-light transition above and below the central axis define the diameter of the inner wall of the capillary.

In the first few frames, while the capillary is heating up, there will be little expansion. When the greatest inner diameter among the tracked sections achieves a database-specified expansion (typically 120% of the original diameter), the maximum diameter is established by parabolically interpolating through the three cuts corresponding to the largest transverse sections. In subsequent frames, all cuts are equally spaced between the half-maximum expansion points, as shown in Fig. 9.

If, during the TRACK operation, the program establishes that the processed output of any given frame is too noisy, which can be caused by glowing dust particles, polyimide remnants, and the like, then that frame is rejected.

The MEASURE Operation. The last operation, MEASURE, applies the diameter-measuring algorithms described for TRACK, above, to a dense family of sections of the cooled bubble, for the purposes of characterizing the bubble and monitoring the performance of the lathe. Derived information includes the maximum diameter of the bubble and its axial position, and the axial positions of the 90% and 10% expansion sections as illustrated in Fig. 10.

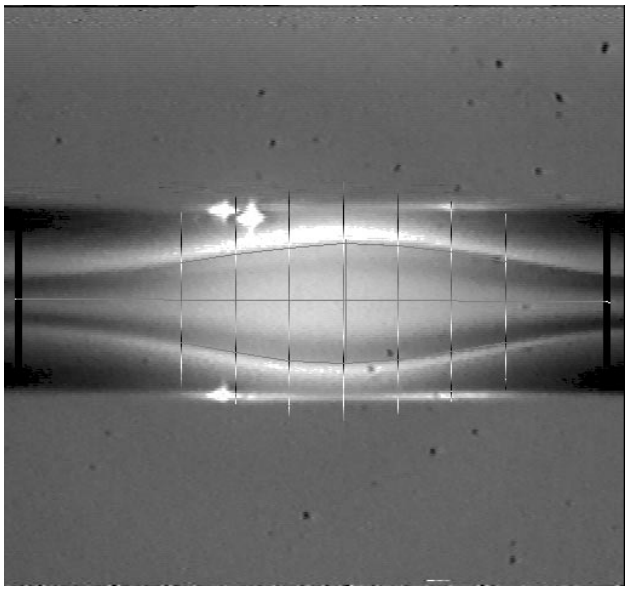


Fig. 9. The TRACK algorithm is used to monitor the capillary inside diameter growth at several axial segments along the forming bubble (the seven vertical lines). Inner edges within each segment of the capillary are tracked in real time. The three largest measurements are fitted with a parabola, the maximum value of which is taken to be the bubble diameter. Cell inflation is progressively arrested as the bubble expands.

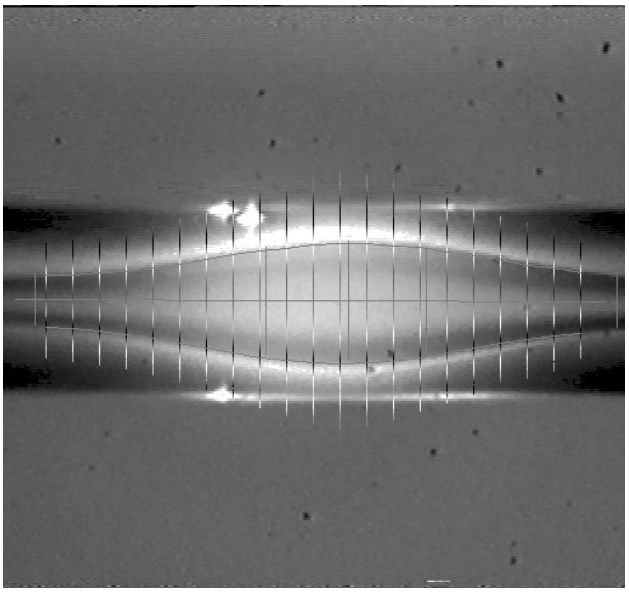


Fig. 10. After blowing is complete, time is less critical and a more complete characterization can be made. The MEASURE operation determines the cell diameter, length, and axial position to an accuracy of 2%.

Conclusion

The geometry of the bubble cell raises capillary electrophoresis to dramatically higher levels of sensitivity and linearity without sacrificing resolution. Application of automation and computer vision to the manufacturing process enables efficient fabrication of consistent, high-quality, customizable, documented bubble cells. Over 40,000 bubble cells have been produced to date.

Acknowledgments

The bubble cell and BubbleWorld efforts were a pleasurable collaboration crossing several organizational and geographical boundaries. The authors are eager to take this opportunity to acknowledge the important contributions of others to the success of this project. During the HP Laboratories EGAD phase (electrophoretic globular absorbance detector), Dick Lacey was responsible for the optics, Jim Young the product design, and Bill Gong the automation electronics. Gratitude is expressed to Barry Willis and Sid Liebes for their continued support, and to Doug McManigill for recognizing the importance of CE to HP. Lastly, special thanks are extended to the Waldbronn Analytical Division's CE business team for so elegantly commercializing this technology.

References

1. M. Albin, Paul D. Grossman, and S.E. Moring, "Sensitivity Enhancement for Capillary Electrophoresis," *Analytical Chemistry*, Vol. 65, no. 10, May 15, 1993, p. 489.

2. *Electrophoresis Capillary with Dispersion-Inhibiting Cross Section*, U.S. Patent #5,324,413 (Gordon), initially filed September 15, 1989.

3. *Rectangular Capillaries for Capillary Electrophoresis*, U.S. Patent #5,092,973 (Zare), filed January 26, 1990.

4. New Product Announcement, R&S Medical, Mountain Lakes, New Jersey.

5. Y. Xue and E.S. Yeung, "Characterization of Band Broadening in Capillary Electrophoresis due to Nonuniform Capillary Geometries," *Analytical Chemistry*, Vol. 66, no. 21, November 1, 1994, pp. 3575-3580.

6. M. Smith, "An Environment for More Easily Programming a Robot," *Proceedings of the IEEE International Conference on Robotics & Automation*, May 1992.

Bibliography

1. *Capillary Zone Electrophoresis Cell System,* U.S. Patent #5,061,361 (Gordon), filed March 6, 1989.

2. G. Gordon, "High-Sensitivity, High-Linearity CE Bubble Cell Detector," poster presentation, *HPCE '91*.

HP Disk Array: Mass Storage Fault Tolerance for PC Servers

In the process of offering a new technology to the marketplace the expertise of the user is often not considered. The HP Disk Array offers RAID technology with special installation and configuration features tailored for ease of use.

by Tom A. Skeie and Michael R. Rusnack

The rapid proliferation of PC-based local area network systems and the growing dependence on them has resulted in concern about the reliability of such systems, particularly in the area of data storage. PC-based systems do not have the history or perception of reliability that minicomputer or mainframe systems have built up over the years. Thus, it is no surprise that concern about fault tolerance is becoming a critical issue in the PC community. This concern has led to a great deal of interest and research into external storage systems and redundancy.

The significant improvement in CPU performance in the 1980s (50 to 100% per year) created the need for similar performance improvements in other system components, such as memory and system storage. Improvements in memory performance came mostly from new memory architectures and algorithms rather than in the memory components. Similarly, magnetic hard disk technology, the most commonly used component for system storage, could not provide the needed performance improvements. A technology called RAID (Redundant Array of Inexpensive Disks)[1] was proposed as a method for improving the I/O throughput for system storage. The basic proposal suggested that the I/O bandwidth could be improved by spreading the data over multiple hard disks to obtain some level of concurrency. However, because of the unacceptable trade-off in system storage reliability caused by this data distribution,[2] multiple data-redundancy algorithms (RAID levels) were proposed. A brief tutorial on RAID technology is given on page 74.

The benefits of improved I/O throughput and data protection have made RAID technology widely accepted on computer platforms of all kinds, including PC servers. While many architectures can efficiently implement a RAID storage system, the generic architecture includes the mechanism for implementing one or more of the RAID algorithms (typically a hard disk controller) and a method for communicating and controlling a group of hard disks (see Fig. 1).

The HP Disk Array implements RAID using a hardware-based EISA-to-SCSI controller in which multiple RAID algorithms are implemented in the controller's firmware. A controller implementing RAID is commonly referred to as a disk array controller. For the HP Disk Array controller five hard disks are connected to each of the controller's two SCSI buses giving a total of 8G bytes (assuming RAID level 5) of protected capacity controlled by each controller. Fig. 2 shows the hardware components of the HP Disk Array. In addition to the benefits of RAID, the HP Disk Array offers a wide range of features, such as predictable failure notification, automatic disk-failure detection, and automatic disk-failure recovery. What really differentiates the HP Disk Array from a majority of similar RAID products is the focus on making the complex RAID technology readily available to PC network administrators who may not have an in-depth knowledge about RAID technology, SCSI technology, or the target operating system.

After an investigation of other systems offering RAID technology, we found that it was difficult to install and configure many of them properly and to recover from a disk failure, even for knowledgeable PC network administrators. We saw this as an opportunity to differentiate the HP Disk Array product from other disk array products. New ease-of-use software utilities were designed specifically to simplify the usability of the HP Disk Array product.

The final product includes several complementary ease-of-use features such as:

- A disk array controller optimized for performance with the selected disks
- Hard disks prejumpered, configured, and mounted in hot-plug* enclosures
- A disk cabinet with automatic SCSI ID selection

* Hot-plug in this context is the ability to replace a disk while the system is in operation.

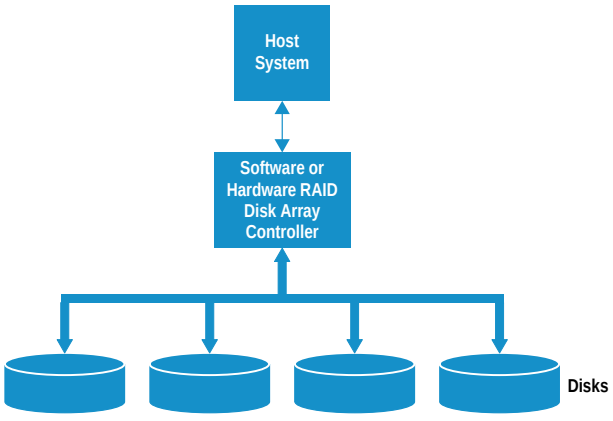


Fig. 1. Generic RAID architecture.

Fig. 2. The components of an HP storage system. The SCSI-II controller, the disk array controller, and the disk module make up the HP Disk Array.

- Software designed specifically for easy configuration and integration of the RAID system into any network operating system.

This article describes the decisions made during the design phase of the HP Disk Array that resulted in making the RAID technology significantly more usable for the typical user without sacrificing valuable product features.

Storage Categories

In a typical computer there are many places to store information either temporarily or permanently. Some of these storage media include RAM, shadow RAM (BIOS), video RAM, read cache, write cache, hard disks, flexible disk, and tape. Whatever the storage medium, computer storage can generally be classified as either primary storage or secondary storage.

Primary storage is often referred to as the system memory, or just the memory. The memory is typically made from semiconductor components and is characterized as being very fast, volatile, and expensive. Computers use memory for immediate and temporary storage. This is the storage containing the data that the CPU operates on and manipulates. Most servers today typically have about 16M bytes to 64M bytes of primary storage.

Secondary storage is often referred to as the system storage, or just the storage. The storage medium is most often magnetic and is slower, nonvolatile, and much cheaper than memory. The storage is therefore used as a permanent repository of information. Servers typically have between 1 Gbyte and 100 Gbytes of storage.

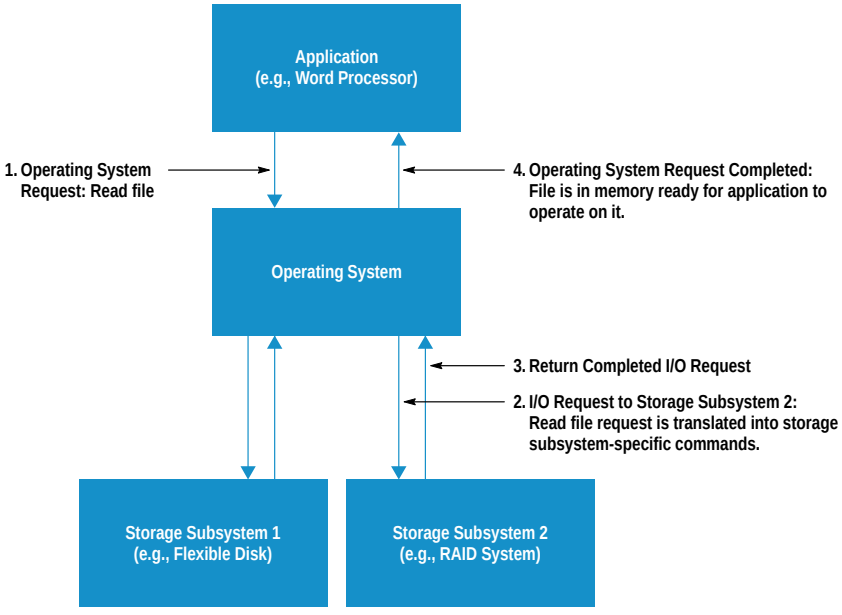


Fig. 3. A typical I/O request flow.

The relationship between these two types of storage is based on how information is transferred in a typical application.

The I/O Process

When an application such as a word processor wants to open a file, it is usually the operating system that is responsible for completing that request. Depending on where the file is currently located (it could already be in memory), the operating system may generate one or more I/O requests to the storage system to bring the file into memory. This process is illustrated in Fig. 3.

A couple of issues should be pointed out here. First, the application does not know how or where the operating system has stored the file it is requesting. Similarly, the operating system may not know how or where the storage subsystem has stored the data making up the file.

For instance, in a RAID system one I/O request from the operating system may result in multiple transfers in the storage subsystem. Since RAID 1 (see page 74) mirrors or duplicates all information from one hard disk to another, every time the operating system requests a file to be written, the RAID subsystem must write this file twice. If the RAID subsystem is using RAID 5, writing a file may result in multiple reads and writes.

Since several I/O transfers may result from a single request from the operating system, something in the system must have the intelligence to distribute the data properly and know how the data should be retrieved. Obviously, in a modular system such as the one described above many approaches to implementing this intelligence in a RAID storage system can be imagined. The following section explores the most common RAID storage system approaches and some of the advantages and disadvantages of the different approaches.

RAID Architectures

A wide variety of architectures have been explored in the effort to provide better I/O throughput and data protection based on RAID technology. These architectures can be divided into two major categories based on whether the RAID algorithms are executed in the host CPU or a dedicated CPU. When using the host CPU, the architecture is generally thought of as a software solution, and when a dedicated CPU is used, the solution is said to have a hardware architecture. The major difference between these architectures most often amounts to a trade-off between cost and performance.

Software RAID Architectures. Since a software RAID architecture uses the host system CPU to execute, one obvious disadvantage with this approach is the potential for lower overall system performance than with a dedicated CPU as in the hardware RAID architecture. However, in many environments, the host CPU is not used to its full capability. Therefore, using the server CPU to execute RAID algorithms provides the most cost-effective RAID architecture while still providing adequate system performance.

Several operating systems now offer some method of data redundancy by implementing the RAID algorithms as part of the operating system (see Fig. 4a). The advantage of this approach is that there is no additional cost to the user. However, the performance may be affected in CPU-intensive application server environments such as database engines.

Not every RAID level is offered through the operating system, and some operating systems offer no data redundancy scheme at all. This has prompted some storage system vendors to provide custom software that implements the desired RAID scheme. Fig. 4b shows the software architecture in which the the desired RAID level is provided with an intelligent device driver.

Hardware RAID Architectures. In environments where the server is used for CPU-intensive applications, adding the overhead of executing RAID algorithms may significantly

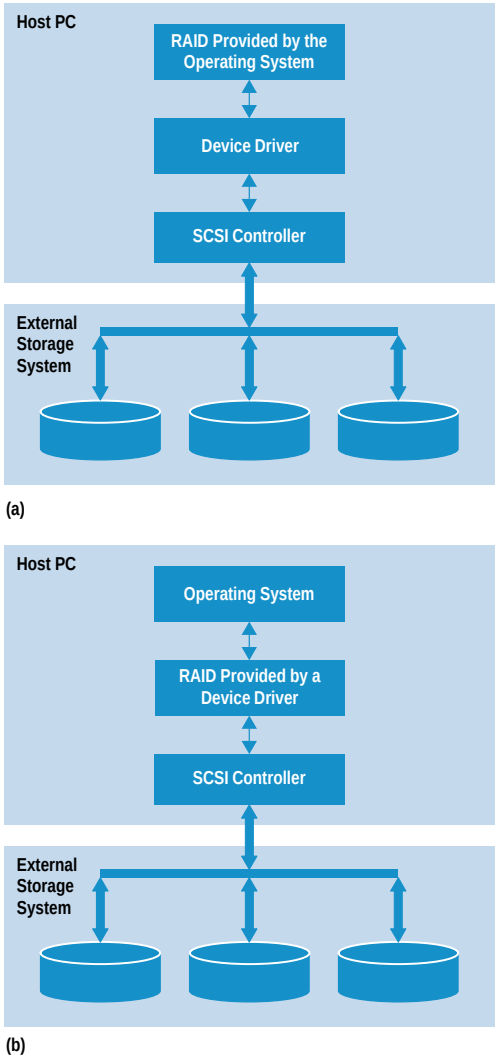


Fig. 4. Software approaches to implementing the RAID technology. (a) Software inside the operating system. (b) Customized software drivers.

An Overview of Raid Technology

The Need for Information Storage

The trend today is to put increasingly larger amounts of an organization's information onto some sort of electronic media for quicker, concurrent, and geographically independent access. The information being put on the electronic media is often critical to the operation of the organization. This trend has created an increasing demand for newer and better methods of electronic storage . One such electronic storage method is data protection via RAID (redundant array of inexpensive disks) technology.

RAID technology finds its roots in a series of papers referred to as the Berkeley papers.[1] Although these papers do not precisely define RAID, they do provide the basis for the architecture of the technology. The need for disk drive redundancy developed as the need for data integrity and system reliability became a growing issue. For example, the consequence of a single point of failure in a disk storage system could result in the failure of the entire computer system. This loss of productivity and often data was deemed unacceptable.

Seven RAID levels are defined. Each level specifies a different disk array configuration and data protection method, and each provides a different level of reliability and performance. Since only a few of these configurations are practical for most online transaction processing systems, file servers, and workstations, RAID levels 0, 1, 3, 5, and 6 are described here.

RAID 0

Although it is often debated that since RAID 0 is not redundant and should therefore not be considered as a RAID mode, nearly all RAID solutions include this mode. RAID 0 distributes the data across all the disks in the disk array configuration (see Fig. 1) Since there is no redundancy, the capacity utilization is 100 percent.

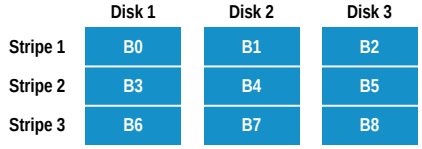


Fig. 1. RAID 0 configuration.

The data blocks in Fig. 1 are broken into three parts and striped across three disks. What this means is that while data is being retrieved on disk 1, the data on disks 2 and 3 can be requested and ready to transfer sooner than if three I/O requests were made.

RAID 1

To answer the need for reliability and data integrity, system managers have often implemented a solution in which write data is mirrored on two separate disk systems. This implementation is referred to as RAID 1. The primary advantage of RAID 1 is its simplicity. RAID 1 provides a slight improvement in read performance over the other implementations. However, write performance is poor because all data is duplicated. The primary disadvantage of RAID 1 is cost, because for every byte of storage used on a system an equal amount of storage must be provided as a mirror. This results in a cost differential of 100% over standard nonredundant mass storage.

In Fig. 2, the data is mirrored on each disk. In the event of a failure on disk 1, the same data is available on disk 2.

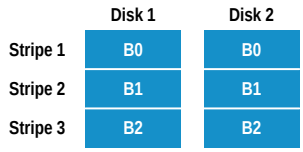


Fig. 2. RAID 1 configuration.

RAID 3

The architecture for RAID 3 is often referred to as a parallel array because of the parallel method that the array controller uses in reading and writing to the disk drives. For RAID 3 it is necessary to provide two or more data disks plus an ECC (error correcting code) disk. Data is dispersed or striped across the data disks, with the ECC disk containing an exclusive-OR of the data from the other disks. Unlike the other RAID solutions, the data is dispersed across the disk in a byte interleave rather than the typical block interleave. With the spindles all synchronized, the data is placed on the same cylinder, head, and sector at the same time. In the case of RAID 3, each drive is connected to a dedicated SCSI channel, which further ensures the performance.

This architecture can handle any single disk failure in the chain. If a data drive fails, data can be recovered from the failed drive by reconstructing the exclusive-OR of the remaining drives and the ECC drive. The advantage of this scheme is that redundancy is achieved at a lower cost (compared to RAID 1). The primary disadvantage is its I/O performance for small amounts of data. When an application requires the transfer of large sequential files, such as graphic images for workstations, this is the best method.

Fig. 3 shows the configuration for RAID 3. The data is striped across drives like the RAID 0 configuration. If a failure occurs, the data is reconstructed based on the parity data on disk 3.

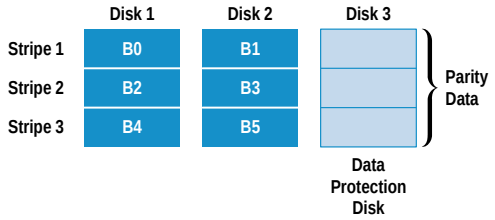


Fig. 3. RAID 3 configuration.

RAID 5

RAID 5 was defined in an effort to improve the write performance of RAID 1 and RAID 3 systems. Like RAID 3, the data blocks are distributed over the disk drives in the system, but unlike RAID 3, the ECC data is also distributed across all the drives (see Fig. 4). With this configuration reads and writes can be performed in parallel.

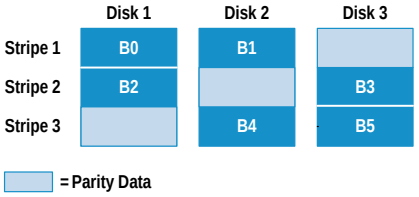


Fig. 4. RAID 5 configuration.

RAID 6

Like RAID 0, RAID 6 is not yet accepted as a standard RAID configuration. This configuration not only mirrors data (like RAID 1) but also stripes the information (see Fig. 5). Because the data is striped, the performance is very similar to that measured in a RAID 0 configuration. The penalty for use of this configuration is that 100 percent more disk space is required.

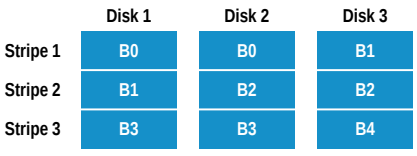


Fig. 5. RAID 6 configuration.

Reference

1. D. Patterson, G. Gibson, and R. Katz, "A Case for Redundant Arrays of Inexpensive Disks (RAID)". *ACM SIGMOD Conference Proceedings*, Chicago Illinois, June 1988, pp. 109-116.

lower the overall system performance. A separate CPU dedicated to executing the RAID algorithms is therefore used in a hardware RAID architecture to offload RAID algorithm execution from the server CPU. Since this architecture requires additional hardware, the major disadvantage here is a higher cost than a software architecture.

The hardware RAID approaches can be categorized as internal or external solutions, referring to whether the RAID hardware resides inside or outside the host server. Internal RAID hardware has to interface directly with a host system bus, such as with an EISA or PCI bus. Disk drives are then connected to the internal RAID hardware directly via one or more SCSI buses (see Fig. 5a). While such a solution may offer superior performance, it does have some subtle disadvantages. Since the RAID hardware must interface directly with the host server hardware and operating system, there is a high level of dependence between the host and the RAID hardware. This dependence affects many aspects of the product's usability, such as the serviceability of failed hardware. For instance, if a problem occurs in the RAID hardware, the entire server must be shut down for repair.

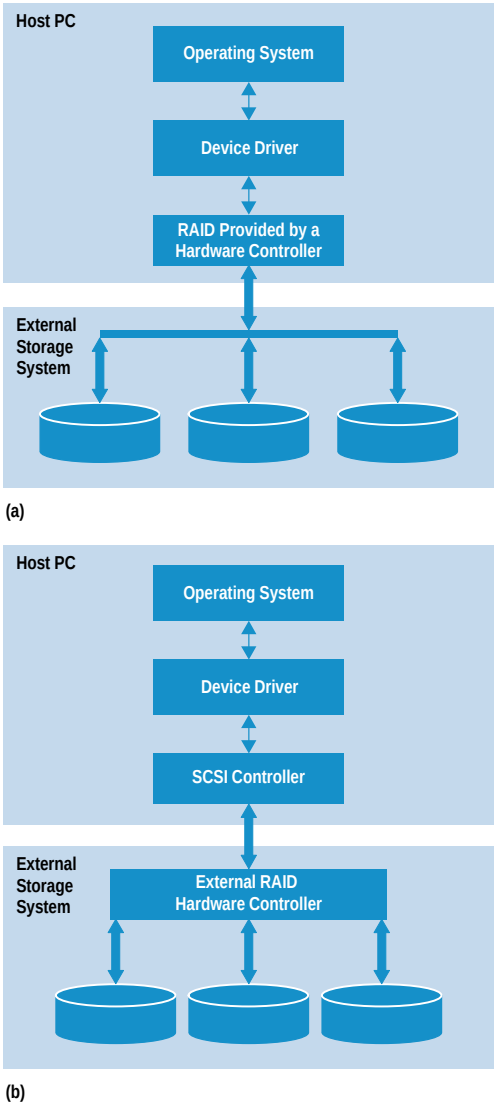


Fig. 5. Hardware approaches to implementing the RAID technology. (a) Inside the host system. (b) External to the host system.

This is not the case with an external RAID hardware solution. Here the RAID hardware is typically residing in the same enclosure as the hard disks, external to the host system (see Fig. 5b) and therefore provides more independence from the host system. In addition to making an external product more serviceable, the independence allows the RAID system manufacturer to more readily support new host hardware systems and new operating systems. Since the external RAID systems require additional hardware, they tend to be more costly and have lower performance than internal solutions.

Many other architectural differences exist in RAID systems that affect the price/performance ratio. For instance, the RAID hardware may use onboard RAM for caching data. Such caching can provide a significant performance improvement, especially when executing RAID level 5 write operations. For a sequential write operation, RAID hardware can calculate the ECC for the stripe without having to perform any read operations if the controller has enough memory to cache one entire stripe. Recall that the amount of data in a stripe depends on two things: the chunk size and the stripe set size. For example, if the chunk size is 64K bytes and there are four disks in the stripe set, the required cache for one stripe would be $4 \times 64K$ bytes = 256K bytes. For the PC server environment, RAID hardware commonly uses between 4M bytes and 64M bytes of onboard RAM shared between read and write operations. Thus, cache can become a significant cost increase to the overall RAID system.

Architectural differences such as the number and type of disk interfaces used (e.g., fast SCSI, fiber, and SSA (Serial Storage Architecture)) can also affect the attributes of a RAID hardware product. For RAID systems that manage large amounts of storage, multiple disk interface buses may be necessary to obtain the best bus utilization. The CPU capabilities will also affect the RAID system price and performance.

The HP Disk Array Architecture

The HP Disk Array product was intended to take advantage of the emerging disk array market for PC servers. It was to be offered as an integrated storage solution for HP's network server product, HP NetServer, as well as a third-party storage product. As discussed briefly above, there are many viable RAID system architectures and features with different end product qualities. Deciding on which architecture and features would best fit the HP Disk Array product was based on the following product requirements:

- Time to market. Because of the dynamics in the PC market, a product (or project) rapidly becomes outdated. If a product is not released within a certain period of time (i.e., "window of opportunity"), it can mean the difference between product success and failure. As mentioned, the HP Disk Array was also to be integrated as part of a PC server developed by a different HP Division. Therefore, two separate programs were dependent on a timely release.

- Operating system support. HP has an established customer base on a wide range of PC server operating systems. While some companies offer new products with limited operating system support, the HP product requirement called for support on all common PC server operating systems. At the time of the first release of the product this included DOS (for boot support), Novell NetWare, IBM OS/2, SCO UNIX®, Microsoft® OS/2, and Banyan VINES.
- Capacity. Based on market research that a typical PC server had a storage need of 1 Gbyte to 5 Gbytes, the HP Disk Array was targeted to offer capacities ranging from 2 Gbytes to 7 Gbytes.
- Performance. Several market research studies found that improving I/O performance was the main reason to purchase a disk array among most PC server users. Performance was therefore an important product requirement.
- Ease of use. Entry-level products should have ease of use as one of their primary attributes. This was not the case with many of the RAID systems we examined. Thus, making it easy for our customers to use RAID technology was an important requirement for the HP Disk Array product.

While all of these product requirements contributed to the product's success, most of them can be considered to be entry-level requirements. These are the criteria necessary to be considered a competitor in this market.

The following sections describe in more detail some of the challenges the R&D team had to deal with to make the HP Disk Array RAID product easy to use.

Installation of a RAID System

Since the PC environment is considered open (that is, hardware and software from one PC manufacturer are expected to operate with any other PC manufacturer's hardware and software), products tend to be very generic when they reach the end user. This often makes the installation and configuration of new hardware and software more troublesome for the user. Installing a new RAID system on a new or existing server is no exception. In fact, it is generally more difficult since the hardware and the RAID system must be configured. The installation of a RAID system involves three major steps: installing and configuring the hardware, configuring the RAID system, and then making the added storage capacity available to the operating system's storage pool by partitioning and adding a file system. This process can be quite intimidating, even for the most advanced user. Since the last installation step is the same (at least very similar) for any new storage, the following sections will focus on the first two steps.

Hardware Installation and Configuration

After unpacking and verifying that all the parts of a disk array system are available, the user must first install and configure the hard disks and then proceed to do the same for the RAID controller.

Installing and Configuring the Hard Disks. Most RAID controllers communicate with the hard disks via the SCSI (Small Computer System Interface) interface protocol. SCSI is a general-purpose interface protocol that allows multiple devices such as disks to communicate in a peer-to-peer fashion on a parallel bus interface. Each device is referenced with a device number, or SCSI ID. If the device numbers are not properly set up (e.g., two devices having the same ID number), the SCSI system will not work.

In addition to setting the SCSI ID, several SCSI interface parameters must also be properly set for each device on the SCSI bus. Among these are settings for synchronous and asynchronous data transfer, data parity check, and maximum data transfer speed. The SCSI devices are configured by using hardware and software switches.

Some RAID systems provide the opportunity to load-balance the I/O transfers between multiple SCSI buses. For instance, if the RAID system has seven hard disks and each disk can offer a sustained transfer rate of 3 Mbytes/s, a single SCSI 8-bit (narrow) bus with a maximum transfer rate of 10 Mbytes/s could be saturated (7×3 Mbytes/s = 21 Mbytes/s for commands and data and 0.66×21 Mbytes/s = 14 Mbytes/s for data assuming a typical 33% SCSI command overhead).* Thus, the user must know the approximate characteristics of the SCSI bus and the disk drives to distribute the disk drives properly.

Before hard disks can be used for storing useful data, they have to be low-level formatted. This is a process in which the magnetic medium of the disk drive is divided into generic fixed-size sectors (typically 512 bytes) where data can be stored.

Installing and Configuring the RAID Controller. Setting up the RAID controller is perhaps the most difficult and definitely the most critical part of the hardware installation. During this process the user has to power down the server, take the cover off the server PC, and install the RAID controller into an empty slot. If not handled properly, static electricity could cause the entire server to malfunction after this operation. Further, if the controller has hardware jumpers, they must be configured before the controller is installed in the server.

Once the system is put back together again, a configuration program, such as the EISAConfig program for an EISA system, is run to assign host system resources to the controller (e.g., IRQ and BIOS addresses). Since the RAID controller has one or more SCSI interfaces in addition to EISA, it is very common to configure the controller's SCSI parameters with the EISAConfig utility.

RAID System Configuration

Once the hardware is installed and configured, the RAID system configuration can begin. To the user this is probably the greatest challenge in the entire installation process because the RAID technology introduces many new terms and concepts such as stripe set size, chunk size, and RAID level. The user must understand how the selection of a parameter will affect the system in terms of data protection, system performance, and capacity utilization.

Fig. 6 shows the dependency relationships between some of the configurable RAID system parameters. Note that Fig. 6 is not meant as a complete picture, but only as an illustration of the complexity involved in determining the values to assign to these parameters. The operating system is the only

* The maximum number of devices on a narrow SCSI bus is eight, and since the disk controller is considered to be one device, a maximum of seven hard disks can be connected to the bus.

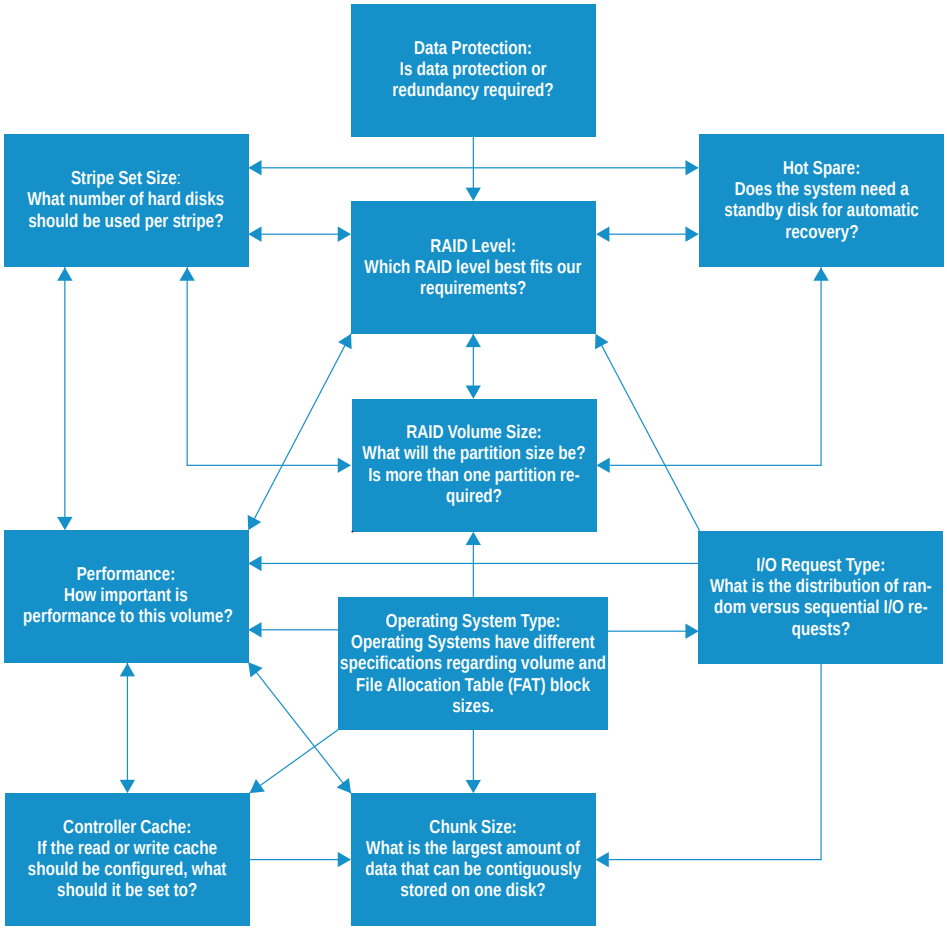


Fig. 6. Dependency relationships between RAID parameters.

entity in Fig. 6 that has no dependencies because its configuration is typically fixed by the time the storage system is installed. Without a thorough understanding of these dependencies, the user will not know what effect the value assigned to one parameter will have on the remaining installation. The following sections look at some of these parameters and their dependencies in more detail.

- Stripe set size. Before any other selections such as logical volume size and RAID level can be set, a decision must be made on how the hard disks should be grouped. A stripe set is the collection of disks that make up an array that implements a RAID level that typically uses a disk striping technique.[3] Disk striping involves spreading data over multiple disks in an interleaved pattern to improve performance (see Fig. 7). Selection of a stripe set greatly affects the remaining configuration parameters such as the RAID level and logical volume size. The behavior of the array will also be affected, especially in terms of its performance. Stripe set size has the following dependencies:
  - Performance. If the hard disk enclosure has multiple SCSI channels with hard disks distributed over multiple channels, it may be desirable to have multiple stripe sets to obtain better performance of the overall system. Also, depending on the nature of the data and the frequency of distribution, multiple stripe sets may be desirable.
  - RAID level. Each RAID level has its own constraints or requirements. For example, RAID level 1 can only be used with an even number of hard disks, while RAID level 5 can be used with an even or odd number of hard disks.

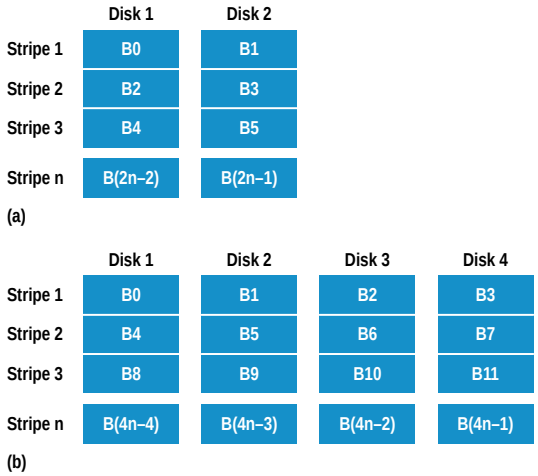


Fig. 7. Two examples of disk striping and different stripe set sizes. Data is spread over disks in equal chunks (or blocks) of data. A chunk is typically between 4K bytes and 64K bytes. By using this scheme of spreading the data among multiple disks, a controller can overlap operations to the disks and thereby improve the overall data access. For example, in (a) two chunks (e.g., B0 and B1) could be read in the same time it takes to read one chunk. If the system bandwidth allows, it is clear that the stripe set in (b) will have more concurrent data access than the stripe set in (a).

- Hot spare. Depending on the number of hard disks available, it may or may not be possible to have a hot spare (standby hard disk) that will automatically replace a failed hard disk in the stripe set. Further, unless all logical volumes defined in the stripe set are configured to be redundant (RAID levels 1, 5, or 6), a hot spare will not serve its intended purpose in the event of a disk failure.

- RAID level. Choosing the RAID level is one of the central parameters in the RAID system installation. It has a lot of interdependencies, including:
  - Data Protection. If the storage system is to withstand a disk failure, some redundancy scheme such as RAID 1 or 5 must be selected. If no protection is required, RAID 0 may be the best choice offering best capacity utilization and good performance.
  - Performance. Selecting a RAID level will have a significant impact on the I/O performance. For instance, if the I/O request types are random writes, RAID 5 is the lowestperformance selection one can make and RAID 6 would be a better choice.
  - RAID volume size. RAID levels 1 and 6 will consume 50% of the available storage capacity, while RAID 5 costs one disk out of the total number of disks in the stripe set. For example, if the stripe set consists of five disks, RAID 5 will offer four disks or 80% of usable capacity for data storage.
- RAID volume size. Some RAID systems allow multiple partitions within one stripe set. This allows the user to customize the storage by having a mix of volume sizes, RAID levels, and caching schemes. Some dependencies for the volume size include:
  - RAID level. As already mentioned, the RAID level dictates what remains of usable capacity.
  - Operating System. Some operating systems, such as OS/2, cannot use partitions larger than 8G bytes. Therefore, it may be necessary to make multiple partitions because of such operating system limitations.
- Chunk size. Selecting the proper amount of contiguous data stored on one disk (chunk size*), is probably the most difficult task in configuring a RAID system. It is easy to explain what chunk size is, but challenging to explain the implications of selecting different chunk sizes. Chunk size mainly affects the performance of the storage system, but the difficulty lies in predicting how the performance will be affected. The following considerations are important in selecting chunk size:
  - Operating system. Based on the I/O request size, the chunk size should be set to some multiple of the request size. For Novell NetWare, this poses an interesting dilemma because the I/O request size can be selected for each volume during installation of the file system. Other operating systems, such as SCO UNIX, have a fixed I/O request size.
  - Controller cache. Cache use will also affect how the chunk size should be set. For example, it is often desirable to fit full stripes in cache for RAID 5 writes.
  - I/O request type. The more sequential the requests, the more benefit larger chunk sizes will have.

The HP Disk Array Configuration and Installation

It is obvious from the discussion above that configuring a RAID system is not a trivial task. This section describes some of the features the HP Disk Array provides to make it simple for users to install and configure a RAID system.

To simplify hardware installation of the HP Disk Array, the hard disks come preformatted and preconfigured from the factory. Since HP is able to control the hard disks used with the RAID controller, it is possible to optimize all configurable parameters at the factory. Any configurable parameter that cannot enhance the product but could instead lead to less than optimized configurations has been removed. Even the SCSI ID selection has been automated. By selecting the

* Chunk size is also referred to as block size.

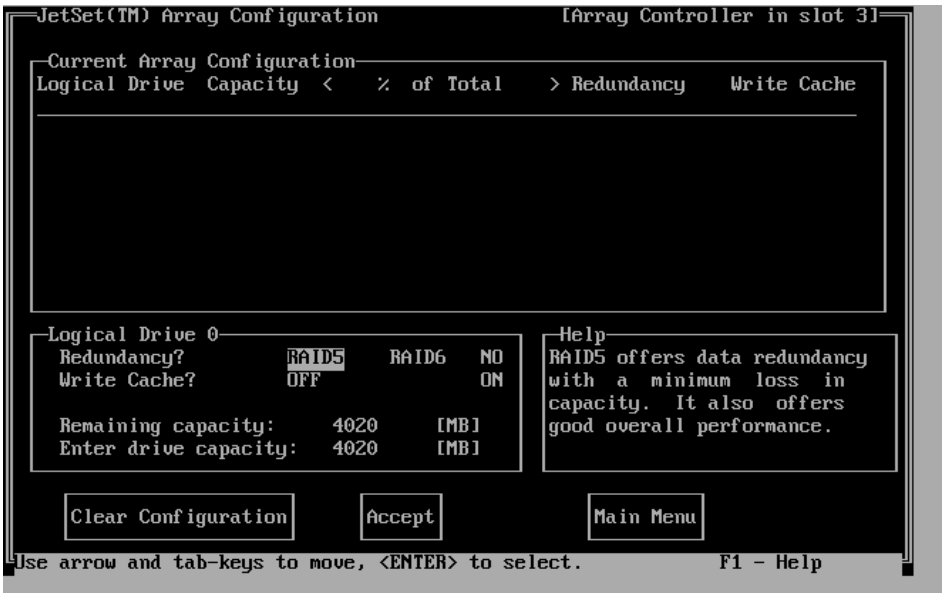


Fig. 8. JetSet configuration screen.

Fig. 9. A RAID configuration in which a DOS partition must be created.

slot in the cabinet, the SCSI ID is also automatically determined. For the RAID controller, the configurable parameters have been optimized as default settings. The SCSI bus use has also been optimized by matching the characteristics of the known disk drives that can be used with the product. The user does not have to understand any performance or configuration aspects of SCSI.

After the hardware has been installed, the user must configure the disk array. A bootable flexible disk is provided that prompts the user to select the language of choice. The installation process will then allow the user to configure the array.

The array configuration utility used with the HP Disk Array is called JetSet. Some of the features of JetSet utility include:

- Sensing and guiding. JetSet will sense if the array is unconfigured, if new hard disks have been added, or if a recovery operation is required. It will then guide the user to perform all the necessary steps to complete the operation.
- Limit choices. JetSet will only present options that make sense for the current operation. For example, a hot spare hard disk does not make sense for a RAID 0 volume, and the option should not even be offered to the user. This approach also limits clutter on the screen.
- Standard defaults. If the user presses the **Enter** key continuously through a few option screens, the array will be configured to the most common defaults. The user has to make different selections to deviate from these defaults.
- Context-sensitive help. JetSet offers extensive context-sensitive help throughout all menus. The goal here is not to require the user to consult a manual during the configuration process.
- Limit new terminology. JetSet is designed to keep the language as simple as possible and not use engineering jargon . A new technology typically brings new terminology. However, many new terms can often be explained with more familiar terms.
- JetSet for all operating systems. To further simplify the learning curve for the user, JetSet is made available in the same form for all supported operating systems. This provides the benefits of simplified software control and user's manual.

A sample of the main JetSet configuration screen is shown in Fig. 8. The user is only prompted to make decisions relevant to the configuration process.

Once the RAID system has been configured and initialized, it is ready for use by the server's operating system for data storage. If the RAID system is the only storage on the server, it can be made bootable by creating a DOS partition and assigning a DOS file system. This is the case for Novell NetWare as shown in Fig. 9.

This is an important point because some RAID systems will not allow the creation of multiple partitions within one stripe set. This increases the cost of the system because the user must either assign a separate hard disk apart from the RAID

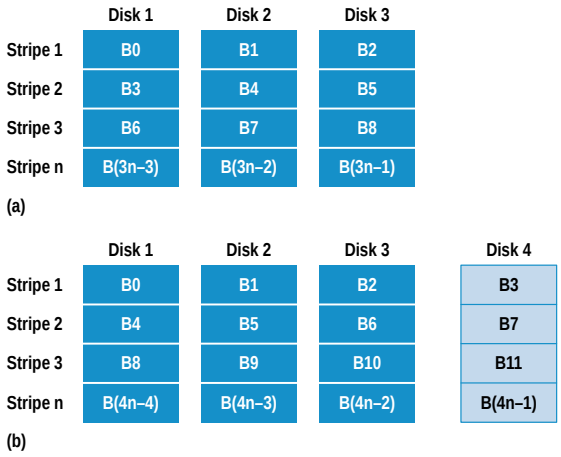


Fig. 10. (a) Data layout in a RAID 0 system. (b) Data layout after adding a new disk.

system as the boot device, or if the RAID system is the only storage system on that server, the DOS bootable partition would be a minimum of a single disk size (1 to 2 Gbytes) and if protected, would take that number times two. The HP Disk Array allows eight volumes for a stripe set using any mix of capacity, RAID level, and caching strategy.

Balancing ease of use and a rich feature set can be very challenging because most manufacturers of products such as RAID systems want to make the products general enough so that they can be used on any host system in any operating system environment with the most features of any available product (a marketing dream). Such generality also challenges the simplicity and ease of use of the system. The HP Disk Array is an example of how few trade-offs were made.

Adding Capacity to a RAID System

Adding capacity to an existing RAID configuration is a well-known dilemma with this technology. Since the data is spread over many disks, a new disk cannot simply be added without shuffling the data around. In Fig. 10 data is spread over three hard disks in an interval of block size B. Each hard disk has n blocks and therefore the array has n stripes. To be able to add capacity to the existing stripe set, some of the data shown in Fig. 10a has to be moved onto the additional disk such that the block layout ends up looking like Fig. 10b. This example is specific for RAID 0.

Adding capacity becomes more complex when the system is configured as a redundant RAID level. Most RAID systems require a complete backup, reconfiguration, and restore for a capacity increase. Algorithms are being developed to allow a capacity increase for an existing stripe set, but are beyond the scope of this discussion.

The HP Disk Array offers a new and simple scheme for adding capacity called *emergency capacity*. Rather than solving the complex problem described above, emergency capacity simply adds a new stripe set to the RAID system (see Fig. 11).

The caveat here is that the new stripe set does not have any disk array characteristics such as added performance and data redundancy. In fact, it is simply a JBOD (just a bunch of disks) configuration. However, it meets the customer's need of immediate capacity expansion. At a more convenient

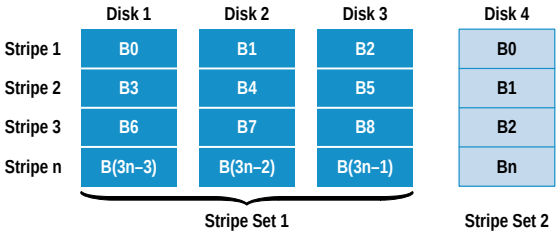


Fig. 11. Adding emergency storage capacity in an HP Disk Array configuration.

time, the user can perform the proper operation for capacity expansion.

Recovering From a Failure

Another major reason why organizations invest in RAID systems is protection against system downtime. According to the March 1994 issue of Byte Magazine, a survey of 450 information service executives at 1000 major companies found computer downtime cost an average of U.S.$78,191 per hour and occurred, on average, nine times a year. A typical outage cost U.S.$300,000 including the cost of recovering or reconstructing the data. However, failures can and do occur, making notification of a failure essential. Some failures do not cause any disruption in the I/O processing (performance may suffer some), while other failures will bring down the entire server.

Failure Notification. Before any manual restoration process can begin the system administrator needs to know that a problem has occurred. The failure notification method is therefore a very important part of the RAID storage system. The notification should be immediate (i.e., within minutes of the failure), visible, and understandable to the administrator.

Many approaches are used for failure notification. Standards such as SNMP (Simple Network Management Protocol) and

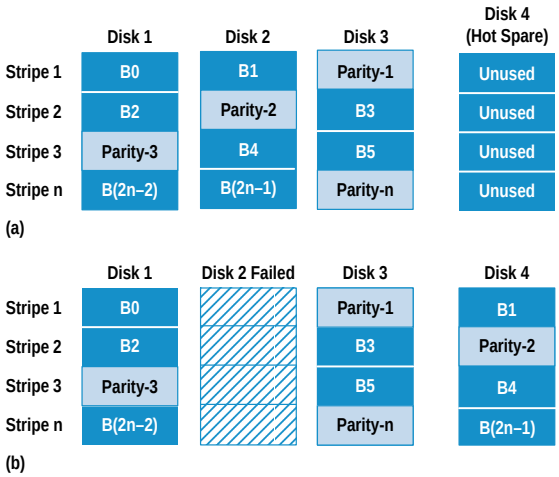


Fig. 12. Failure recovery. (a) RAID system with a hot spare before the failure. (b) After disk 2 has failed and the hot spare is brought online.

DMI (Desktop Management Interface) are emerging. However, since standards have a tendency to take a long time before they are complete enough for the industry to use, industry tends to take its own direction.

Failure Recovery. As mentioned, some failures are recoverable. One such recoverable failure is a single hard disk that is part of a redundant RAID stripe set. The storage system may have the ability to recover by itself by rebuilding on a hot spare (standby hard disk). The failed disk should be replaced to make it the new hot spare. If the system does not have a hot spare, the failed disk must be replaced before rebuilding can take place.

Fig. 12a shows the storage configuration before a failure and Fig. 12b shows what happens after a failure on disk 2 and the completion of using the hot spare (disk 4) to rebuild.

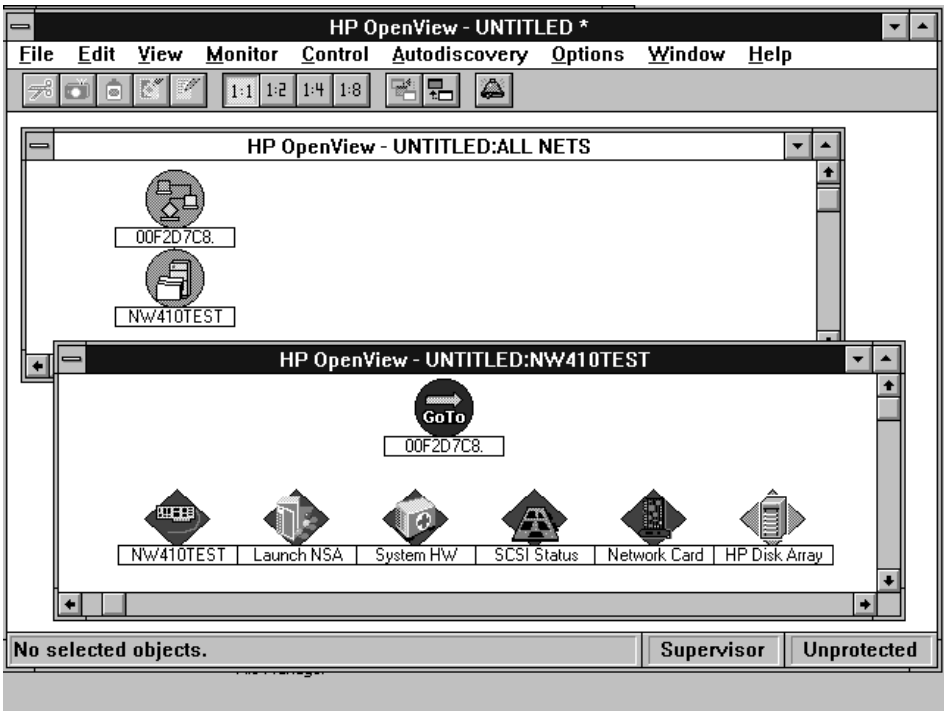


Fig. 13. HP NetServer Assistant/ OpenView screen for remote servicing.

The HP Disk Array takes failure notification one step further. In addition to offering notification of failures that have already happened, it offers a predictive failure notification so that for some failures, the user will be warned about failures that are about to happen before the system is degraded.

It is essential to get notified about failures at some accessible place. If the server is not physically close to where the system administrator is located, audible alarms and messages on the server are not very useful. The HP Disk Array offers an integrated warning mechanism using SNMP traps to an HP NetServer Assistant/OpenView remote console. This allows the system administrator to view error logs and failures from any location (see Fig. 13).

The HP Disk Array offers automatic failure detection and restoration of redundancy with the use of a hot spare disk drive. If no hot spare is used, the online version of the JetSet utility is required to start the restoration process. Upon loading, the online JetSet utility will notify the user about which disk has failed and needs to be replaced before the restoration can begin.

Conclusion

In developing products based on emerging technologies, it is the responsibility of the designers to make sure that these technologies are quickly accepted and available to the typical user. However, in the effort to bring a product to market in a timely manner, the end user is often not prioritized. The result is a product that reflects the complexity of the technology it is built on.

The overall goal for the HP Disk Array design team was to provide RAID technology to the average PC network administrator. Defining what an average PC network administrator meant and what such a user knew about RAID technology took center stage throughout the development. The resulting product is easy enough to use for a novice user, but still provides options and information for the more advanced user.

References

1. D. Patterson, G. Gibson, and R. Katz, "A Case for Redundant Arrays of Inexpensive Disks (RAID)". *ACM SIGMOD Conference Proceedings*, Chicago, Illinois, June 1988, pp. 109-116

2. M. Schulze, G. Gibson, R. Katz, and D. Patterson, "How Reliable Is a RAID?," *COMPCON Spring 1989*, San Francisco, California, February 1989.

3. *The RAIDBook: A Source Book for RAID Technology*, The RAID Advisory Board, Lino Lakes, Minnesota, 1993.

COBOL SoftBench: An Open Integrated CASE Environment

With the aid of a mouse and a menu-driven interface, COBOL programmers new to the UNIX® operating system can improve their productivity with a tightly integrated toolset that includes an editor, compiler, debugger, profiler, and other software development tools.

by Cheryl Carmichael

Many large companies today are still developing mission critical COBOL applications, and they want tools to help them easily transition their developers to open systems.

In 1991, we recognized that many of our customers were moving MIS applications to the UNIX operating system using C SoftBench and C++ SoftBench. These customers asked for a more business-oriented language—specifically COBOL. In addition, HP-UX* had just been recognized as the leading UNIX system environment for MIS organizations moving from mainframes to open systems.

We heard customer concerns about moving MIS development teams to open systems. They wanted to minimize the loss of developer productivity associated with learning a new operating system and new development languages and rewriting their custom applications.

The COBOL SoftBench family is based on HP MF COBOL, Hewlett-Packard's implementation of Micro Focus COBOL, which is based on technology from Micro Focus, Ltd. COBOL SoftBench leverages a development team's COBOL expertise and protects its investment in COBOL applications.

When we were designing COBOL SoftBench we aimed at novice UNIX-system users. These users minimize their learning curves with the aid of a mouse and a menu-driven interface. Several tools have easy-to-use graphical user interfaces, and common UNIX commands are accessible via system menus.

COBOL SoftBench improves developer productivity with a tightly integrated toolset that includes an editor, compiler, performance analyzer, debugger/animator, static analyzer, file comparator, email, and more. Other items provided by COBOL SoftBench to improve productivity include:

- Graphical views of complex programs
- Dozens of integrated third party tools
- Provisions for developers to integrate their own tools using the SoftBench encapsulator[1]
- The ability to automate repetitive tasks using the SoftBench message connector[2]
- Support for mixed-language development with COBOL/C SoftBench and COBOL/C++ SoftBench.

COBOL SoftBench

Our tightly integrated toolset increases developer productivity. The interactions between the SoftBench programs that make up this toolset are shown in Fig. 1.

- The SoftBench program editor† edits program files, shows the column and line location of the cursor, accesses version control for a file, and performs common UNIX commands on a file.
- The SoftBench program builder† controls compilation and linking of application files, provides views and jumps to source code for compilation errors, creates makefiles, and displays makefile dependency relationships graphically. The program builder also supports embedded CICS (Customer Information Control System) statements and SQL applications.
- The SoftBench COBOL animator uses an enhanced window interface to the Micro Focus COBOL animator to debug

† See reference 3 for more information about these SoftBench tools.

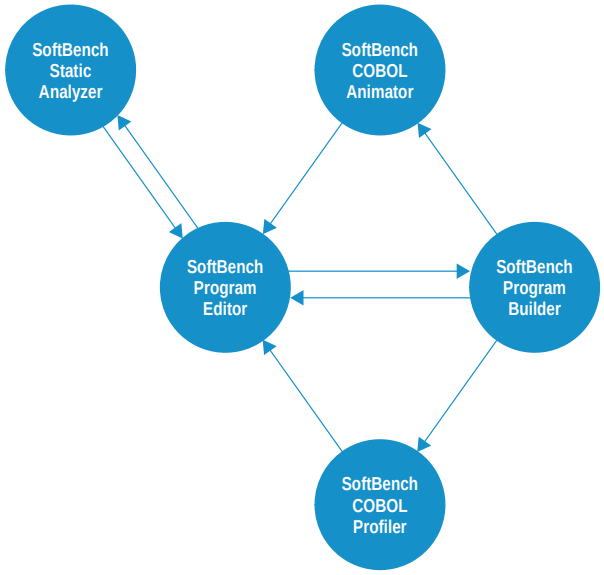


Fig. 1. Interactions between SoftBench tools that make up the COBOL SoftBench toolset.

executable COBOL programs, provide access to native commands for expert Micro Focus users, display data monitors, and allow mixed-language debugging.

- The SoftBench static analyzer enhances visual code browsing by graphical or textual representation of program structure.
- The SoftBench COBOL profiler uses an enhanced window interface to the Micro Focus COBOL profiler to supply detailed statistics about the run-time performance of a COBOL program. It determines performance bottlenecks and helps the developer make changes for the greatest improvements.

Other integrated tools that are not shown in Fig. 1 include:

- The SoftBench development manager† handles file and directory management and version control.
- Computer-based training comes with each SoftBench product. Each computer-based tutorial tells about individual tools and exposes the user to a wide variety of SoftBench capabilities using several hands-on training examples.

Friendly User Interface

COBOL SoftBench developers are buffered from the standard UNIX command line entry format. They select buttons, toggle options, choose menus, and drag-select text. Typing is left for data entry and editing source code. The following model shows how a user would view the UNIX operating system through the COBOL SoftBench interface.

The SoftBench development manager directory list presents a list of files in the currently viewed directory and allows various actions to be performed on any file. For example select Directory: Show Full Listing to view a complete directory listing including permissions, owner, group, size, and modification time (see Fig. 2). The files displayed in the directory list can be limited by using predefined or custom filters.

Within the directory listing, a single line, multiple lines, or even discontiguous lines can be selected for the same action. To select discontiguous files, hold down the **Ctrl** key while clicking the left mouse button on each file name. For example, to compile scattered source files, select the discontiguous .cbl files shown in Fig. 3 and then choose Actions and then one of the compile selections.

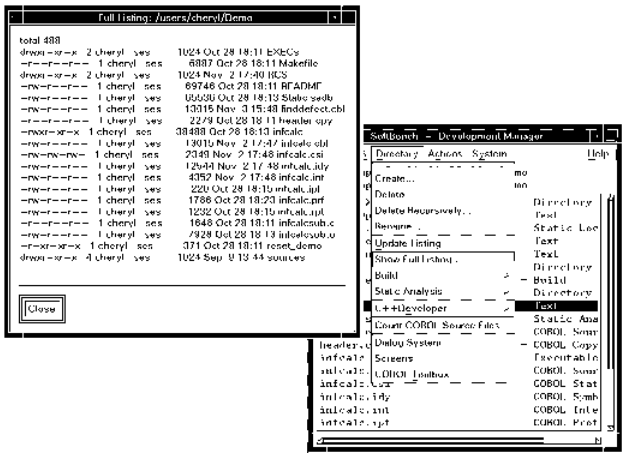


Fig. 2. A directory listing produced by using the SoftBench development manager's Directory: Show Full Listing menu item.

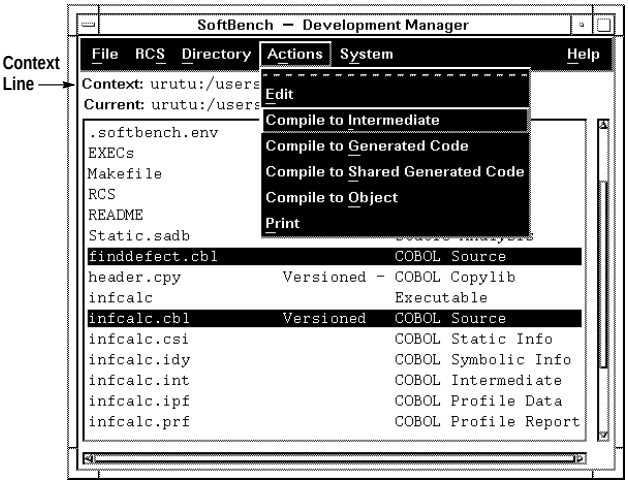


Fig. 3. Selecting several COBOL files for compilation.

Moving about in a directory only requires a mouse movement. To change to another directory and view its contents, move the mouse focus to the context line. While holding down the right mouse button, move through directories until the desired directory is highlighted. Release the mouse button and a new directory list displays.

The SoftBench development manager helps manage files and directories on the developer's computer and across the network. From the File: menu item, the developer can easily choose Edit, Copy, Rename, Import, or Delete depending upon the task.

A number of commonly used UNIX commands are available from the System menu. These include:

- Changing a file's permission
- Printing a file with a number of printer options
- Checking the spelling in a file
- Counting lines, words, and characters in a file
- Creating a backup copy of a file
- Updating the modification time for a file.

Supporting Code Use Model

New COBOL SoftBench users want to leverage their COBOL development expertise and protect their investment in existing COBOL applications. The following example shows how a COBOL developer who is maintaining unfamiliar code would use COBOL SoftBench to tackle a problem that might arise in a COBOL software maintenance environment.

In this scenario the developer receives a complaint that the inflation calculation program (infcalc.cbl) takes too long to start. To deal with this complaint, the developer must quickly identify the performance bottleneck, fix the problem, and verify that the fix does indeed improve the applications performance.

Source Code Control. The inflation calculation program has a fileset containing infc,alcsub.c, infcalc.cbl, header.cpy, and Makefile. The developer sets up a working directory using the SoftBench development manager and checks these files out of a revision control system. The revision control systems that COBOL SoftBench can be configured to use include RCS, SCCS, PVCS, ClearCase, and others.

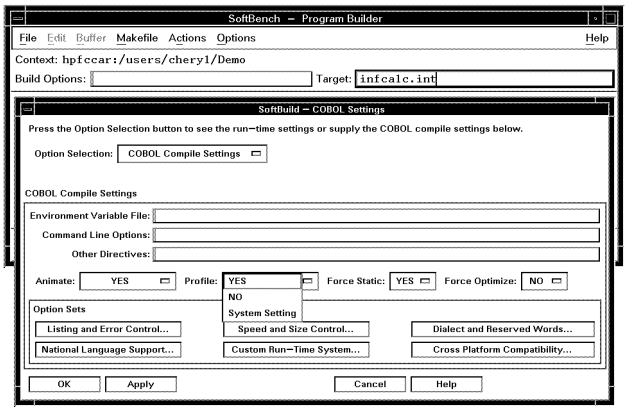


Fig. 4. The SoftBench program builder and the COBOL compile settings dialog box.

Compile Options. The application must be compiled to enable the generation of performance statistics. Using the SoftBench program builder, the developer enables the Profile option with a mouse click in the COBOL compile settings dialog box and compiles the application (Fig. 4). The COBOL SoftBench environment accepts many control options which tailor the compile for special needs.

Micro Focus Run-Time Settings. Over 200 options are available in the Micro Focus COBOL development environment. COBOL SoftBench provides a way to view and change options and values easily. Over 50 of the most common options are available from the SoftBench program builder via the COBOL compiler settings and COBOL run-time settings dialog boxes.

Application Execution. The application is executed to generate performance data. When the developer runs the application from the SoftBench program builder, the files infcalc.rpt and inflcalc.ipf are created.

Application Performance Analysis. The SoftBench COBOL profiler makes it easy to determine where the application is spending its time. Remember that for our example, the performance complaint is about slow startup. The developer notices that a large amount of time is being spent in the initialization module (Fig. 5).

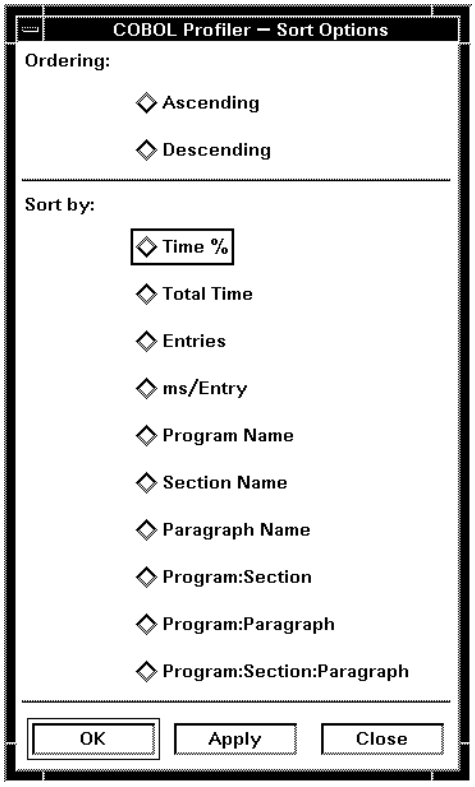


Fig. 6. The ten options for sorting performance statistics. In this example the modules in the program being evaluated will be sorted according to the amount of time spent in each module.

The SoftBench COBOL profiler is an encapsulation of the standard Micro Focus COBOL profiler tool. The COBOL SoftBench implementation enhances the Micro Focus profiler listings of performance data by displaying the data in a SoftBench window. Flexible reporting capabilities are available. For example, there are ten ways to sort performance statistics (see Fig. 6). Using the Options: Display Options from the profiler window, the developer can also choose to view only entered data, nonentered data, or all the paragraphs in the program.

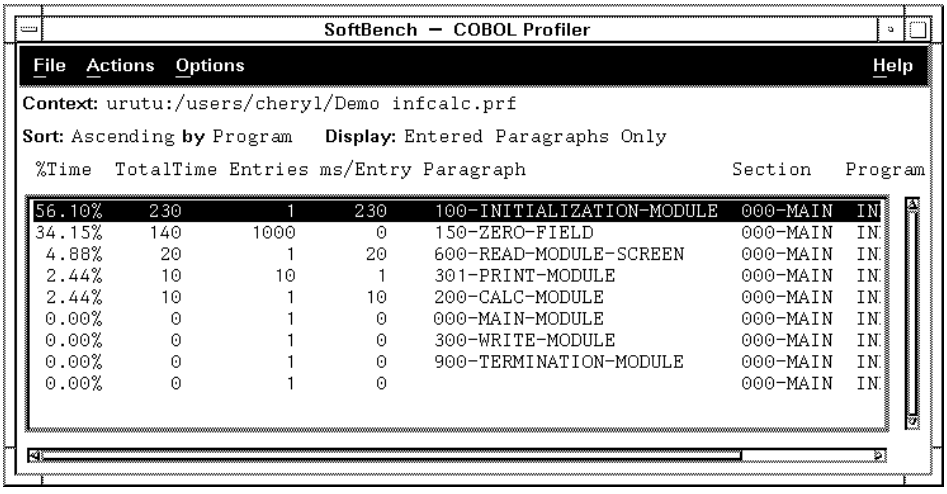


Fig. 5. The profile statistics for the calculation program.

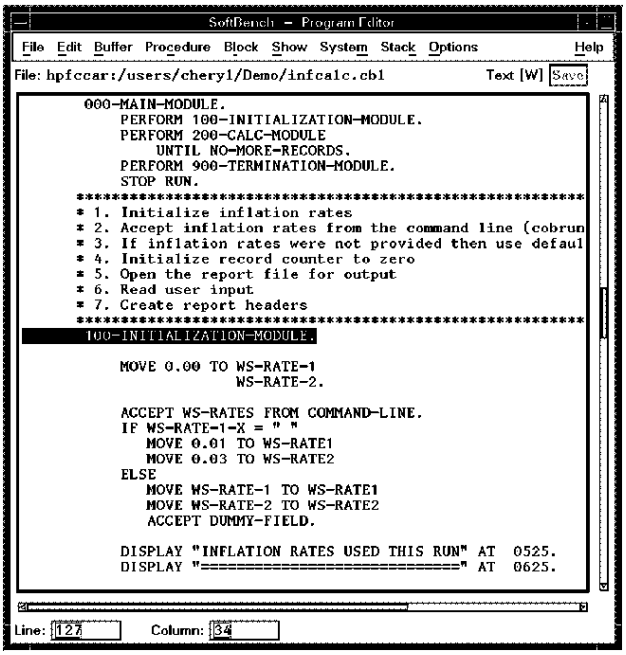


Fig. 7. The Softbench program editor showing the initialization module.

Program Structure. The SoftBench static analyzer allows the developer to save time in searching for program errors in unfamiliar COBOL code. From the SoftBench COBOL profiler, the developer double clicks on the line in Fig. 5 showing the initialization module. This action brings up the SoftBench program editor displaying the source code, starting at the beginning of the selected module (Fig. 7).

The developer needs to learn about the relationships between the many parts that make up the application without having to read the unfamiliar code. Using the mouse button, the developer drag-selects 000-MAIN-MODULE and chooses the Show: Definition( ) menu item to bring up the SoftBench static analyzer. Now, the developer can quickly display the program structure graphically using the Graph: Query Graph menu item (Fig. 8).

Existing Relationships. The static query graph shows that the initialization module performs the 150-ZERO-FIELD paragraph. Looking back at the SoftBench COBOL profiler in Fig. 5, the developer sees that this paragraph is called 1000 times!

Within a very short time, the developer knows where the program is spending time, has an understanding of the program structure, and is ready to debug at a specific paragraph. In other words, the developer has a good start toward solving the problem.

Program Animation. Programmers frequently need to understand a program's logic. By dynamically watching lines of source code execute, they can identify where the behavior of the program differs from what they intended.

Using the SoftBench COBOL animator, the developer selects the line of code that performs the paragraph in question, PERFORM 150-ZERO-FIELD UNTIL ZERO-COUNTER = 1000 (Fig. 9). Choosing Breakpoints: Zoom executes the program through the step just before the perform statement in question. While the animator is paused, the developer can view the code that is

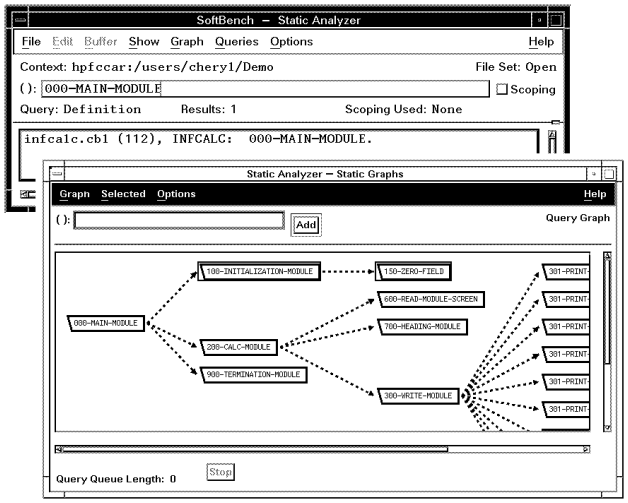


Fig. 8. The SoftBench static analyzer and the static graph showing the structure of the calculation program.

about to be executed. Looking at this paragraph reveals that the WS-ARRAY array elements are being set to zero.

The developer now selects the Animate button. The animate mode single steps through the program. As each line executes, watch windows show the name and value of all variables that change. The developer finds that the program executes this loop 1000 times, just as it was intended.

Textual Static Queries. The Show menu shown in Fig. 10 helps locate any place a specified identifier is referenced. From the SoftBench COBOL animator, the developer drag-selects the WS-ARRAY identifier and then chooses Show: References( ). There are only two references: the declaration and the perform statement.

A more experienced developer could have avoided the animator steps by editing the 150-ZERO-FIELD node from the static

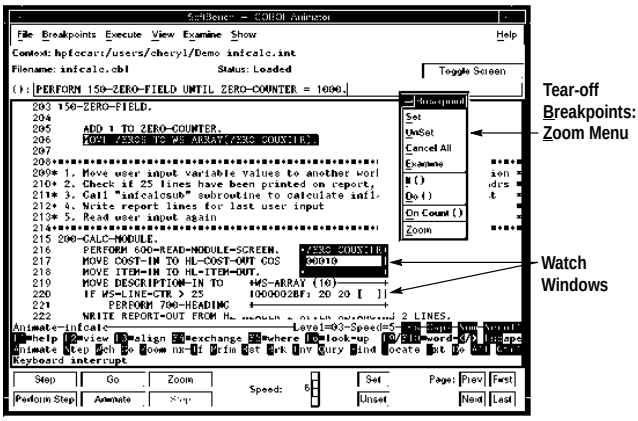


Fig. 9. The SoftBench COBOL animator showing a tear-off Breakpoints: Zoom menu and two examples of watch windows. A tear-off menu is a group of menu commands that have been "torn off" from their parent menu. The parent menu in this case is Breakpoints. A tear-off menu allows the user to keep on the workspace frequently used menu items even after the parent menu item is no longer selected. Fig. 3 shows case in which a menu could become a tear-off menu. By clicking on the dashed line, the menu items shown below the Actions command would become a tear-off menu.

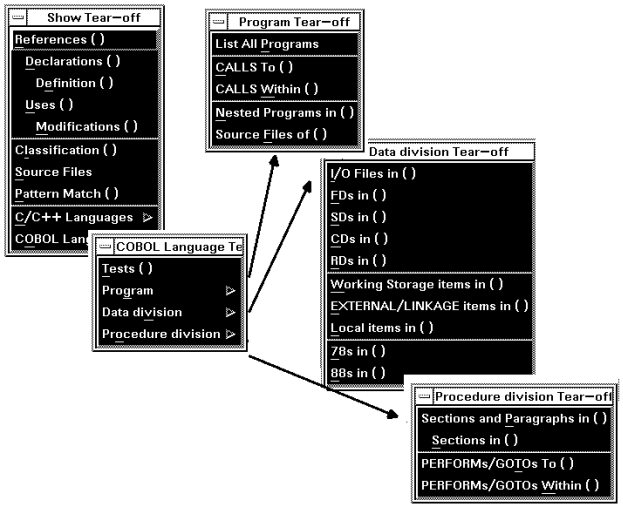


Fig. 10. The Show menu is used to help locate where a specified identifier is referenced in a program.

graph and checking the references for WS-ARRAY from the SoftBench program editor.

Source Code Modification. This defect fix is a simple one: delete the unnecessary perform loop. From the SoftBench COBOL animator, the developer chooses File: Edit and edits infcalc.cbl. Then from the SoftBench program editor, the developer searches for the ZERO-FIELD string, removes the unused code, and saves the changes to the file.

If the developer had needed to make major changes to the code, the SoftBench program builder would take away the pain of finding compile errors. After the developer saves and compiles the program, any errors found are displayed in the Build Output area. To browse and fix errors, the developer can select the First button or select any error in the list. The associated source file appears in the SoftBench program builder, with the text cursor located at the beginning of the line where the error was detected (Fig. 11).

Improved Application Performance. Before the code is returned to the version control system and production, it is important to confirm that a lower percentage of time is spent in the initialization module. The developer recompiles the program, with the Profile option still enabled, runs the program again and from the SoftBench COBOL profiler, selects the Entered Paragraphs Only display option. This shows that the 150-ZERO-FIELD paragraph is not entered during the application's execution (Fig. 12).

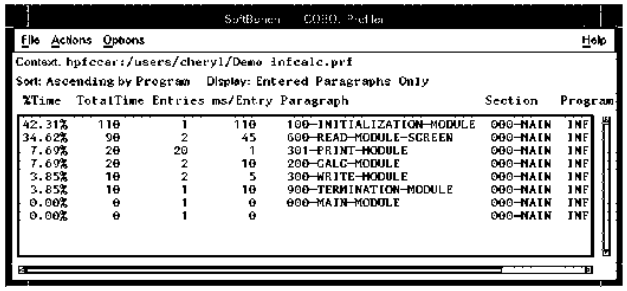


Fig. 12. Profile of the calculator program after it is fixed.

Mixed-Language Model

Development teams can take advantage of other technologies with COBOL SoftBench. For example, consider the developer who is writing a mixed-language application. One program module is written in COBOL, while another is a subroutine written in C. The COBOL module calls the C subroutine.

To illustrate the COBOL SoftBench model, consider the scenario in which the developer has set up a working directory using the SoftBench development manager and is using version control on the the two source files mixedtest.cbl and cpart.c. The program is not working as the developer expects. To help determine what is wrong with the program, the SoftBench COBOL animator and the SoftBench program debugger (for the C subroutine) are used.

Custom Run-Time System. With a mixed-language application, a custom run-time system is part of the application. From the SoftBench program builder, the developer creates a makefile, which detects mixed source code files. With this makefile, COBOL SoftBench will automatically create a custom run-time system. Next, the developer sets compile options to set up both programs for debugging and then builds the application.

To verify that the makefile is working as expected, the developer brings up the dependency graph browser (Fig. 13). This graph shows cpart.c is compiled to an object file and mixedtest.cbl is compiled to an intermediate target.

Mixed-Language Debugging. Since the COBOL program is the entry point in this example, the developer starts from the SoftBench COBOL animator, and the mixedtest.cbl source code is displayed (Fig. 14). The developer chooses File: Soft-debug Adopt to display the SoftBench program debugger window. After setting a breakpoint at the C procedure entry,

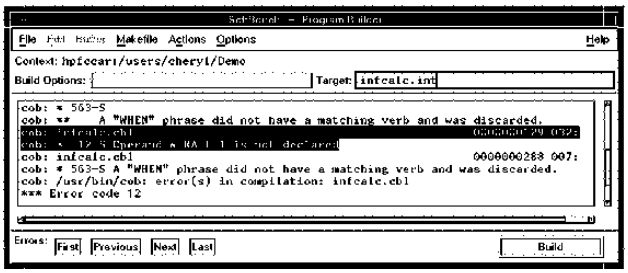


Fig. 11. The SoftBench program builder showing the line in the program where the error occurred.

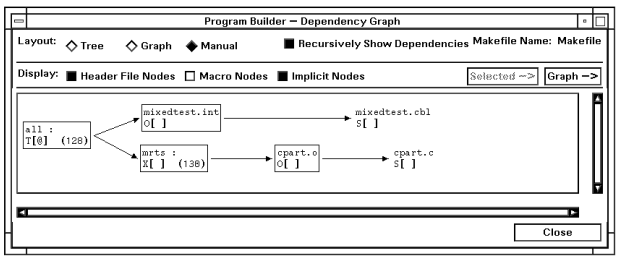


Fig. 13. Dependency graph to verify that the makefile is correctly identifying a mixed-language file.

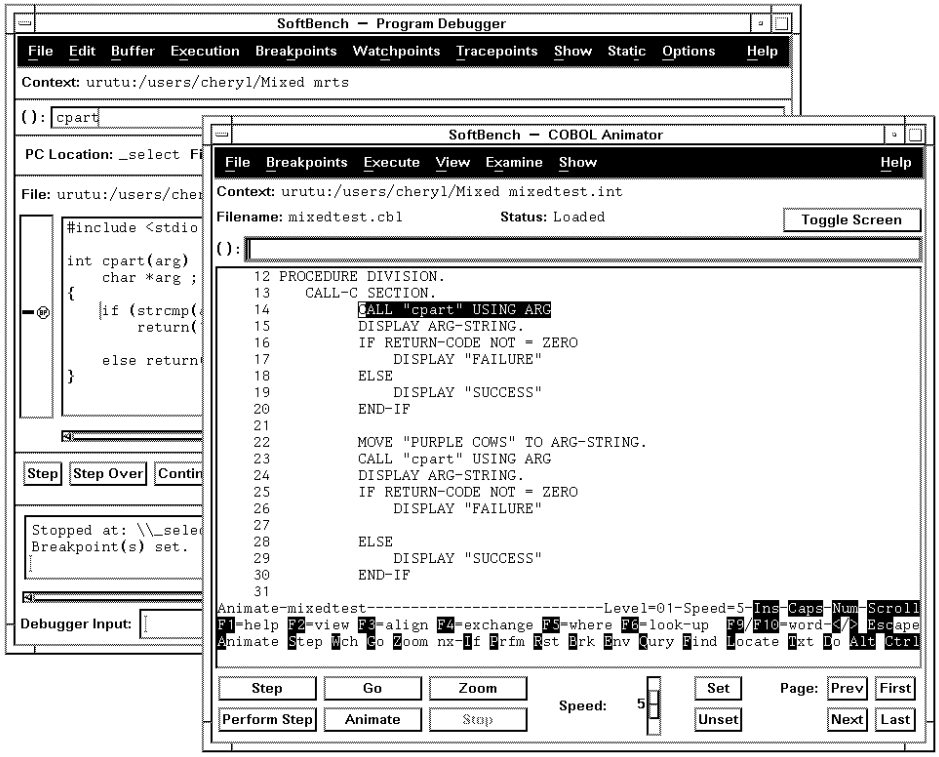


Fig. 14. The SoftBench COBOL animator and the SoftBench program debugger being used to debug a mixed-language program.

the developer selects the Continue button, which returns the control back to the SoftBench COBOL animator.

With both debuggers started, the developer is ready to debug the application using both the SoftBench COBOL animator and the SoftBench program debugger (collectively called debuggers).

Control transfers between the debuggers as the application executes the program module or the subroutine. The developer can apply all the features of either debugger, such as setting breakpoints, watching variables, or setting variable values.

COBOL and C++ Mixed Program. When a custom run-time system includes C++, the C++ language requires that the entry point be the C++ main function. The developer starts the SoftBench program debugger and the main.c source code is displayed. Choosing File: Animate Adopt displays the SoftBench COBOL animator window. The program begins execution at the entry point displayed in the SoftBench program debugger window, and again control passes between the debuggers, depending upon the application's behavior.

Conclusion

HP's integrated application development toolsets, COBOL SoftBench, COBOL/C SoftBench, and COBOL/C++ SoftBench help COBOL programming teams easily transition to open systems. In addition, these products help position the teams to transition to new application development technologies.

Acknowledgments

The COBOL SoftBench products involved many people. Special thanks go to our COBOL SoftBench product team managers Lee Huffman and Dan Magenheimer, our product marketing engineer Pat Hafford, my learning products team Margee Daggett and Tom Huibregtse, and our R&D team Lynnet Bannion, Paul Faust, Jay Geertsen, Robert Heckendorn, Mike Stabnow, Wade Satterfield, and lead engineer Alan Meyer. Additional gratitude to our partners at HP's General Systems Division and California Language Labs, and the Software Engineering Systems Division's core SoftBench team.

References

1. B.D. Fromme, "HP Encapsulator: Bridging the Generation Gap," *Hewlett-Packard Journal,* Vol. 41, no. 3, June 1990, pp. 59-68.
2. J. J. Courant, "SoftBench Message Connector: Customizing Software Development Tool Interactions," *Hewlett-Packard Journal,* Vol. 41, no. 3, June 1994, pp. 34-39.
3. C. Gerety, "A New Generation of Software Development Tools," *Hewlett-Packard Journal,* Vol. 41, no. 3, June 1990, pp. 48-58.

Development and Use of Electronic Schematic Capture in the Specification and Simulation of a Structured-Custom ASIC

ASIC designers must sometimes provide the ASIC vendor with documentation describing the data path of the chip and its relationship to the control portion. This paper describes a method and attendant tools that facilitate the employment of commonly available electronic schematic capture software to ensure that the documentation given to the ASIC vendor always matches the Verilog HDL descriptions used by the ASIC designers for simulation.

by David A. Burgoon

This paper briefly recounts the development and use of schematic capture in the design of a structured-custom ASIC (application-specific integrated circuit). The ASIC was developed through a partnership of two teams of engineers, one from our laboratory and one from the ASIC vendor, which in this case was the HP Integrated Circuits Business Division. The team from our laboratory was responsible for defining the overall architecture of the chip, that is, the overall hierarchical block diagram consisting of data path and control, and the functional description of each module. The vendor team's main contribution was custom design of the data path and implementation of the ASIC's infrastructure, such as pads, scan chain, JTAG† circuitry, clocking network, and the like. Our team was responsible for conveying the logical architecture and functional description of the chip to the vendor team, and was also responsible for verifying the functional design through (primarily) Verilog simulation. The vendor team was responsible for turning the Verilog HDL (hardware description language) functional descriptions into an equivalent physical design.

It soon became clear that if this division of labor was going to succeed, the quality and accuracy of the documentation used to convey the chip's functional behavior to the vendor had to be ensured. Most of our designers preferred to document the overall architecture graphically, and had often done so by drawing block diagrams with their favorite graphics editor. However, to verify the functional design through simulation, Verilog HDL descriptions have to be developed. These descriptions are also used for synthesis of the chip's control logic via the Synopsys toolset (from Synopsys, Inc.). This presented an apparent dilemma: the best way to convey the high-level functional design to the vendor was through schematic graphics, but the only way to describe the functionality precisely was through textual HDL models. If our designers attempted to maintain both forms of description, we faced duplication of effort (essentially describing the block diagram twice, once graphically and once textually), and the possibility of differences between the functional schematic block diagram given to the vendor and the Verilog HDL used for simulation (see Fig. 1).

Goals and Tactics

One way to solve this dilemma would be to develop tools to compare the topology represented by the schematic documentation to that of the Verilog models, and manually reconcile any differences. Another approach would be to try to convert the Verilog models into schematic diagrams programmatically. This approach was discarded because it entailed a lack of control over the aesthetic form of the drawings. The goal we ended up pursuing was to find a way to simulate the documentation, that is, to remove the possibility of disparity between the schematic documentation and the Verilog HDL.

We chose three tactics to achieve this goal:

- Use electronic schematic capture to produce the documentation and the Verilog HDL models. Develop a set of tools and processes that ensure that the hierarchical block diagrams and the Verilog netlists are products of one and the same database.
- Design the tools and processes to enforce a policy that prevents the extracted Verilog text files from being independently altered. All changes to the Verilog models must be effected through the schematics so that disparities are rigorously eliminated.

† JTAG is the Joint Test Action Group, which developed IEEE standard 1149.1, *IEEE Test Access Port and Boundary-Scan Architecture.*

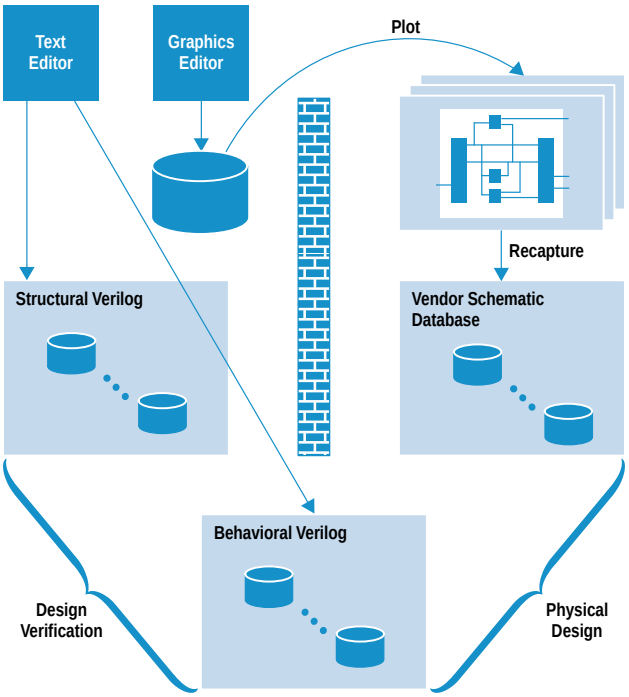


Fig. 1. The designers' dilemma: the block diagram had to be described twice, once graphically and once textually, and there was a possibility of differences between the block diagram given to the vendor and the Verilog description used for simulation. The brick wall symbolizes the lack of connection between the two forms of description.

- Design the tools and processes to be integrated smoothly with the existing simulation and regression testing environment. This allows the extracted Verilog models to be easily verified alongside the handwritten models.

Major Components of the Solution

Our goals were achieved by knitting together tools we had "lying around the house" by means of some tool enhancements, bug fixes, and shell scripts. This approach was in keeping with the grass-roots nature of the effort: we had neither the time, the funds, nor the inclination to search the commercial marketplace for a solution. We were committed to choosing the best tools from among those currently available in our laboratory and integrating them into a solution. Some pieces of the solution were internal and some were commercial. The major components of the solution are described in the following paragraphs (see Fig. 2).

Design Capture System. We chose the Design Capture System (DCS), version 5.10, from Zuken (formerly from HP) as our schematic capture tool. Some of the attributes of DCS that made it an attractive choice are:

- It is a true hierarchical schematic capture tool, not merely a graphics editor that has been coerced into capturing ASIC topology. It implements hierarchy naturally, and has a rich set of built-in features that promote circuit consistency and discourage wiring errors.
- All of our designers were proficient at using it because it had been the tool of choice for many years for the capture of printed circuit board designs.
- A Verilog extraction tool, dcs2ver, had already been written for it.

dcs2ver. DCS comes with a powerful language called DDL (Design Database Language) for accessing the database representing a schematic. Several years ago, a DDL program called dcs2ver was written by our productivity group to facilitate the functional verification of hierarchical printed circuit board schematics through Verilog simulation.[1] It was a natural choice for our application.

Briefly, dcs2ver traverses the hierarchy of the schematic and produces a Verilog module definition for each unique symbol (called a design) in the schematic. Each module definition contains the required port declarations, as well as module instantiations representing all instantiations of symbols on the circuit pages of the design. The result is a set of files, one module declaration per file, that represents the hierarchy of the ASIC as a component-oriented netlist expressed in Verilog HDL.

NgleTalkII. NgleTalkII is a name given to our laboratory's latest generation of simulation, configuration, debugging, and regression testing tools.[2,3] It is listed here because it facilitated the regular testing of dcs2ver-extracted Verilog modules against a mature suite of test scripts written in the

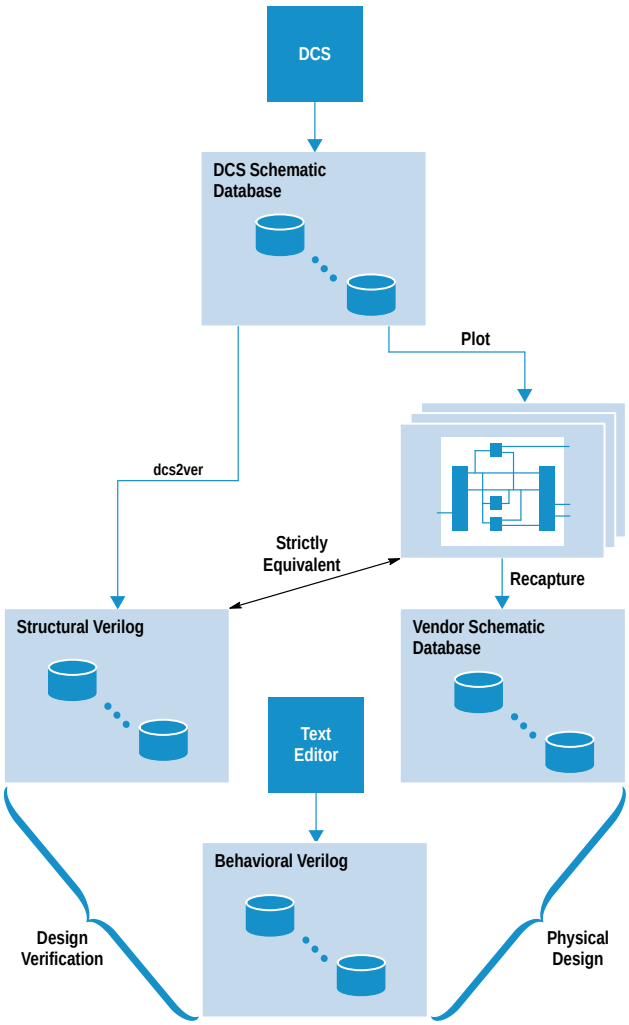


Fig. 2. The solution forces the graphical and textual descriptions to be equivalent.

NgleTalkII language. In other words, NgleTalkII made it easy to "simulate the documentation" on a daily basis.

History Management System. The History Management System (HMS), the internal predecessor of HP's SoftBench CM product, is a set of tools that manage versioned files in a networked environment.[4] It provides mutually exclusive edit access and revision control through a set of client commands reminiscent of the standard RCS (Revision Control System) commands of the UNIX® operating system.

The consistent and universal use of HMS by our design team on all design files made it easy to enforce the policy of disallowing independent modification of the DCS files and the derived Verilog files. This was achieved simply by checking in (placing under HMS control) the original DCS database files, and purposely not checking in the dcs2ver-extracted text files. We used HMS to help ensure that the Verilog embodied in the DCS schematics was extracted strictly from those schematics. Since the Verilog files were absent from the HMS server, the only way to get them (and be sure they were correct) was to play by the rules and use dcs2ver. Thus a typical design change to the high-level structural Verilog required locking the necessary DCS files, making the necessary graphical edits via DCS, checking in the modified DCS files, and running dcs2ver.

Glue and Enhancements

The major components of our solution were integrated through a set of "glue" programs and procedures and enhancements to the existing components.

dcs2ver Enhancements. As mentioned previously, dcs2ver was originally written for board simulation. Our use exposed several weaknesses and bugs, which we fixed. Some of the more notable enhancements were:

- The ability to allow DCS net aliases and splitters with alias labels.†
- The ability to use DCS net "bundles" (heterogeneously named buses),† and to control whether a given bundle is translated to a Verilog concatenation or a set of scalar port connections.
- Control over whether a module is instantiated with ports connected by position or by name.
- Detection of duplicate module names.

HMS Enhancements. As mentioned above, HMS was a key component in our solution. Unfortunately, at the time we were developing our solution, HMS did not "version" (keep old revisions of) non-ASCII files. Not willing to lose this capability for our DCS files, we wrote a set of Korn shell scripts, called rawfci, rawfco, rawfutil, and rawfhist, on top of the similarly named standard HMS client commands. These scripts allowed the DCS files to be fully versioned in a manner that is transparent to the user. This enhancement saw wide use for other types of non-ASCII files, and has since been incorporated into standard HMS.

† Net aliases are a DCS construct; they allow a net to be referred to by another name, an alias. Splitters with alias labels are a means of aliasing net names when splitting off elements of buses. Bundles are like buses, except that each element of a bundle can have a name that is not related to any other element.

make netlists. Not all users of our dcs2ver-extracted Verilog files had the training, license, or inclination to run DCS to extract the netlist files. To accommodate these users, scripts were developed whose invocation was initiated from a make(1) command. This approach was congruent with the existing NgleTalkII environment, in which users were accustomed to performing the sequence

```
fupdate; make; vsim
```

which causes new copies of out-of-date files to be fetched, C-language models to be compiled if necessary, and a Verilog simulation to be started. We added a netlists target to the makefiles which was referenced by the default target.

Briefly, the netlists target conditionally causes a shell script named extract_netfiles to be started which runs dcs2ver under the terminal form of DCS, called DDAS (Design Database Access System). If the user has a license to run DDAS, dcs2ver is executed locally; otherwise, an HP Task Broker[5] job is submitted to a DDAS server. In either case, the needed dcs2ver-extracted files are deposited on the user's machine in a few minutes.

Observations

Our approach to the use of schematic capture in the design of this ASIC was a tolerant one: our designers were free to use DCS as much or as little as desired. The typical designer did not capture a DCS design for every module in the hierarchy of the ASIC, but only went a few levels below the top level and then switched over to hand-generated textual Verilog for those modules that were predominantly behavioral as opposed to structural. One designer had a depth of zero (as measured from the top level) in the DCS hierarchy. Another used DCS only for the data path, leaving the control to a text editor, vi(1). The other two designers fully embraced the experiment, having a depth of five in some places.

Our use of DCS for high-level hierarchical capture was generally successful at meeting our goals. The most significant complaint was directed at our strict enforcement of the policy of not checking extracted netlists into HMS. The path to get extracted Verilog was somewhat complex in its implementation, especially for those who relied on the remote DDAS service. We had some occasional downtime because of licensing and HP Task Broker administration problems. In retrospect, it probably would have been more reasonable to check the extracted Verilog into HMS as nonversioned raw files, and use a cron(1m) job to do a daily make netlists to keep them up to date, thereby also ensuring that any direct textual edits were obliterated.

Another flaw in our approach was that it entailed a duplication of schematic capture effort. The vendor designers normally captured schematics in their environment from textual Verilog supplied by us. In the case of this ASIC, they found themselves capturing schematics whose top levels were essentially identical to the corresponding DCS schematics. As a result, we are already taking the next obvious evolutionary step: our designers are now directly capturing top-level schematics using the vendor's environment and are sharing files with the vendor via HMS.

Acknowledgments

Our ASIC design team consisted chiefly of Frank Bennett, Larry Mahoney, Bryan Prouty, and the author. The vendor team from the Integrated Circuits Business Division (ICBD) consisted of John Pessetto, Brian Miller, and John Morgan. The design team would like to thank our ICBD partners as well as our peers for having patience with us while we developed our tools and processes for schematic capture. We would also like to thank Tim Carlson, the original author of dcs2ver, and all those who enhanced the tool significantly, especially George Robbert.

References

1. L. Mahoney, "GTD Board and ASIC Simulation Tools: The New Generation," *Proceedings of the 1991 HP Design Technology Conference,* 1991, pp. 519-524.

2. R. Jayavant, *NgleTalkII Users Guide, Version 1.5,* internal HP document, May 27, 1993.

3. R. Jayavant, *NgleTalkII Environment Reference, Version 1.0,* internal HP document, May 27, 1993.

4. S.A. Kramer, "History Management System," *Proceedings of the Third International Workshop on Software Configuration Management (SCM3),* June 14, 1991, page 140.

5. T.P. Graf, et al, "HP Task Broker: A Tool for Distributing Computational Tasks," *Hewlett-Packard Journal,* Vol. 44, no. 4, August 1993, pp. 15-22.

Design and Development of a 120-MHz Bus Interface Block Using Standard Cells and Automatic Place and Route Tools

The RW_IO block runs at 120 MHz and interfaces the master memory controller chip's 60-MHz core with the 120-MHz processor bus drivers. A design approach using standard cells, automatic place and route tools, and a powerful database management and build tool was used to construct the RW_IO block. This approach was chosen over a full custom or data-path solution because of its reduced risk and the flexibility of the design tools.

by Robert E. Ryan

The master memory controller chip is one of three ASICs (application-specific integrated circuits) that make up the memory subsystem of certain HP workstations and business servers. The chip interfaces with the processor via the processor bus, which runs at 120 MHz. It also connects to the other memory subsystem components via the memory bus, which runs at 60 MHz. Most of the internal logic of the chip runs at 60 MHz. The RW_IO block, a portion of the chip's logic that runs at 120 MHz, transfers data between the processor bus drivers and the 60-MHz core logic.

Functional Description

The RW_IO high-speed interface block is designed to transfer data and control between the processor bus running at 120 MHz and the core logic of the master memory controller chip, which runs at 60 MHz. To transfer data from the processor bus efficiently requires twice the bandwidth at the 60-MHz interface. In operation, two cycles of the 120-MHz data and control signals are registered and then the registered information is transferred on the next rising edge of the 60-MHz clock. Fig. 1 is a schematic diagram of the RW_IO input logic.

The data transfer is synchronized by delaying the 60-MHz clock by 4.166 ns with respect to the rising edge of the 120-MHz clock (Fig.2). This synchronization scheme requires that the register-to-register transfer of data from the 120-MHz to the 60-MHz clock domains occur within 4.166 ns worst case. The reverse principle is used for transferring data from the core of the master memory controller chip out onto the processor bus. Data and control information is transferred from the 60-MHz to the 120-MHz domains within the RW_IO block. This information is then multiplexed at the 120-MHz rate and presented to the processor bus drivers.

A special case exists when the processor bus is idle and we wish to present the data directly from the 60-MHz core to the processor bus without registering and multiplexing at the 120-MHz rate. A special path (fast path) exists to handle this case. This path also has the requirement of transferring data from the 60-MHz to the 120-MHz domains within 4.166 ns. Fig. 3 illustrates the timing of the fast path, and Fig. 4 is a schematic diagram of the RW_IO output logic.

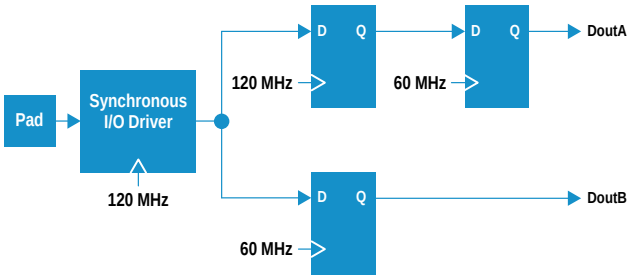


Fig. 1. Input logic of the RW_IO block.

Design Architecture

Five distinct functions are required to transfer data between the processor bus and the core logic. With these five

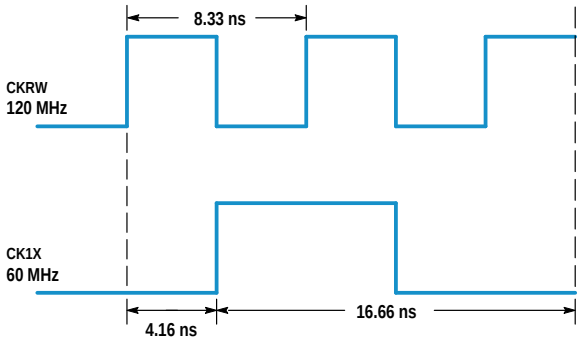


Fig. 2. Relationship of the 120-MHz clock to the 60-MHz clock.

basic functions defined, a bit slice design approach was taken to implement the entire functionality. A bit slice macro was built for each of the five functions using standard cells and Cell3 place and route tools from Cadence Design Systems, Inc. The bit slices were stacked together horizontally to create the complete RW_IO interface block. The following table defines the five bit slice functions and their macro names.

Function	Name
Input only, 2 bits	RWI2BS
Input only, 3 bits	RWI3BS
Bidirectional transfer with fast path	RWIOBS
Bidirectional transfer with fast path and processor bus output gated with data_valid	RWIOVBS
Bidirectional transfer with fast path and multiplexing for address and data	RWADIOBS

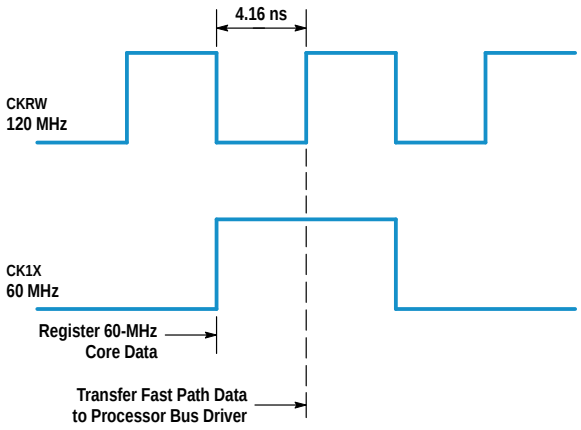


Fig. 3. Fast path timing.

Physical Bit Slice Construction

Because each of the five functions required a limited number of design elements, a semicustom implementation using standard cells was chosen. The most complicated function, RWADIOBS, required 19 cells. The five functions were placed and routed using a proprietary tool called Autopr[1] and Cadence Cell3. Autopr was used to generate the bit slice bounding box and standard cell rows and to place the I/O ports. Special care was taken in cell placement and the routing of the clock signals. Scan control signals and power and ground signals were routed so that connections could be made by abutment when the bit slices were stacked together. Once the construction of the bit slices was complete, Cell3 abstracts were generated for each bit slice using the Cadence Opus framework tools. These abstracts were then used to build up the complete RW_IO functionality.

Physical Block Assembly

The top-level Verilog netlist for the RW_IO block was generated by a script in such a manner that the individual bit slice functions were declared in the order which they physically appear from left to right within the RW_IO block. As part of the netlist generation a stackup file was created which defined the order of the bit slice elements, the pins on each bit slice corresponding to ports on the top and bottom of the block, and the direction of each pin (in or out). A simple awk program was developed that read the stackup file and created two command files for Cell3. The first file directed the placement of the pins on the block to line up with the appropriate bit slice. The second command file directed Cell3 to place the bit slice components in the order defined

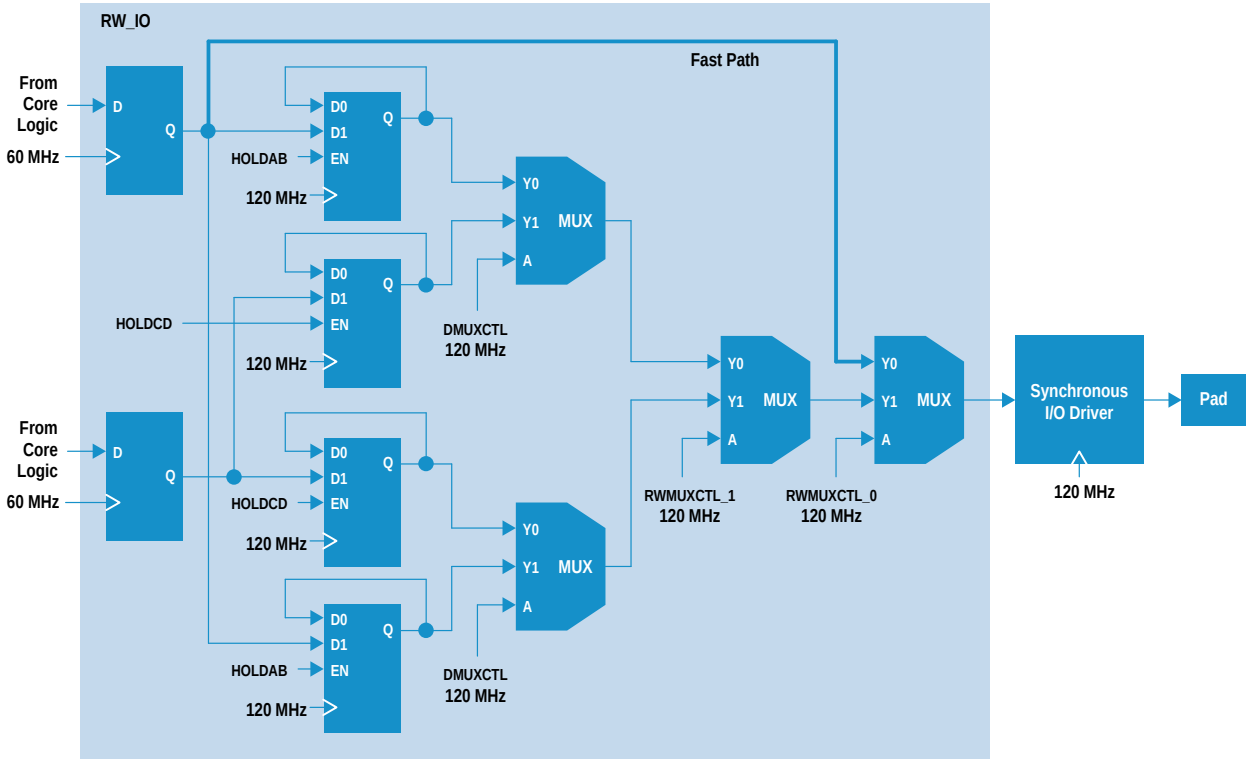


Fig. 4. Output logic of the RW_IO block (RWADIOBS function).

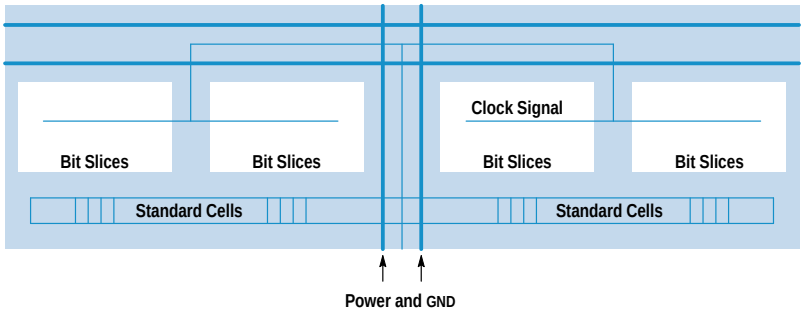


Fig. 5. RW_IO block layout.

by the stackup file. The RW_IO block was separated into four sections to allow the main power, ground, and clock signals to pass through the center of the block and to allow for clock tree taps (see Fig. 5).

The block has a single standard cell row along the bottom where clock buffers and a small amount of combinational logic reside. Cell3's clock tree synthesis capabilities were used to insert buffers in the global data control signals (hold_abL, hold_cdL, mux_ctrl_rw_0, mux_ctrl_rw_1, mux_ctrl_rdata, return_byte_swap). The routing of clock, V_{dd}, and GND signals was customized using specific Cell3 commands.

Characterization

In the early stages of development each bit slice function was characterized separately using Aida's Timver static timing analysis tool, with lumped capacitance and distributed RC delay information back-annotated from Cell3. Timing libraries were generated automatically by Timver for each function. A total of eight Timver runs were required to characterize all the functional paths in each bit slice. The results of the eight analysis runs were merged into a single timing model of the bit slice. A new Verilog netlist was generated from the physical Cell3 database that included the buffers inserted by Cell3. The libraries and the new Verilog netlist were used for the initial master memory controller chip top-level timing analysis. Final top-level timing analysis used a Verilog netlist back-annotated from Cell3 and full distributed RC back-annotation for the block, instead of the Timver-generated timing libraries. This reduced the potential for error during final timing analysis.

Metal Migration Analysis and Power Bus Sizing

The RW_IO block consists of 87 bit slices, 68 of which are the RWADIOBS design. Because the RWADIOBS is the most complex macro and makes up most of the RW_IO block, it was assumed that all 87 bit slices were of this type for the metal migration analysis. The first step in the analysis was to determine the current required by each element in the RWADIOBS macro. This was done by first calculating the amount of capacitance being switched by each input of every cell and then the capacitance being switched by each output, which was the sum of the device output drivers and the wire load (back-annotated from the layout). Once this was done, the equation $I_{ave} = CVf$ was used to determine the current switching of each element. The current requirements of the elements were summed to determine the total for the RWADIOBS block. This value was used to determine the total current required by the RW_IO block and to size the power and ground buses appropriately.

Verification and Extraction

The fully placed and routed block-level database was brought into the Cadence Opus framework where a GDSII and CDL† netlist of the complete RW_IO block was generated. This data was then transferred to Mentor Graphics'* CheckMate tool where layout-versus-schematic, design rule, and electrical rule checks were performed. Checkmate was instructed to flatten the database to the standard cell level and then perform a complete distributed RC extract. This data along with the back-annotated netlist was then used for final full-chip static timing analysis.

Build and Design Data Management

The entire design process, from the construction of the bit slices to final analysis of the RW_IO block was managed by the Atria ClearCase tools. ClearCase is designed for managing large software development projects. The ClearCase tools provided a development environment with configuration management, revision control, and build management. ClearMake, ClearCase's make-compatible build tool, was used to build all the parts of the RW_IO block automatically.

† GDSII and CDL are industry-standard formats for data interchange.

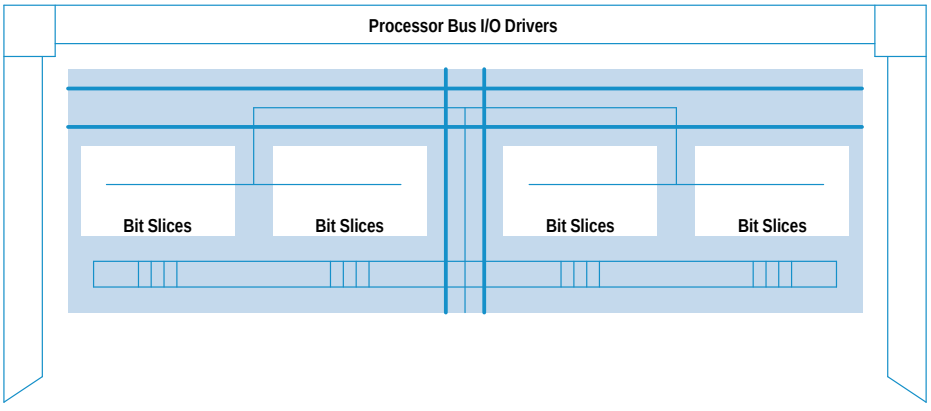


Fig. 6. First version of the RW_IO block.

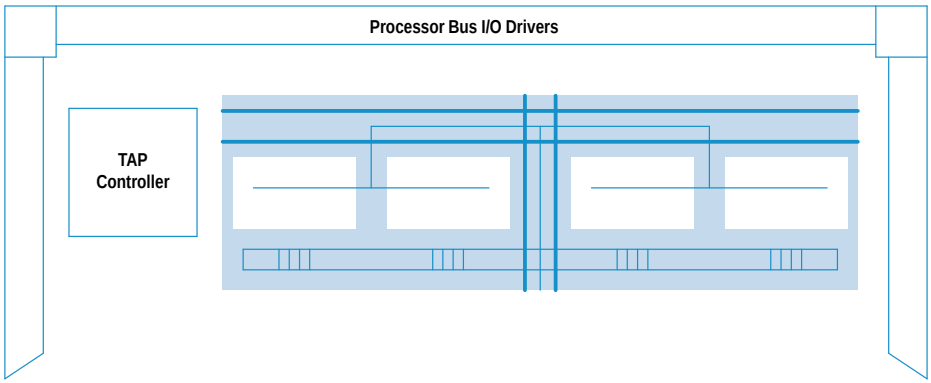


Fig. 7. RW_IO block redesigned for area conservation and inclusion of little endian functionality.

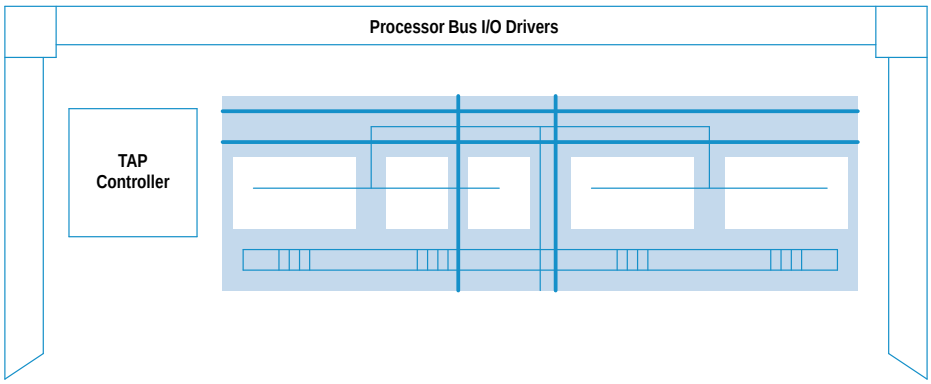


Fig. 8. RW_IO block with relocated power rails.

Makefiles were generated to optimize the overall construction process and to allow for efficient rebuilding if necessary. ClearMake's parallel distributed building capabilities greatly improved the efficiency of the build process by parallelizing the construction of the independent sections and distributing the build processes across the networked machines. During all build steps ClearMake performs auditing of the build processes. ClearMake keeps track of all the files read and written during the build and creates an audit trail called a configuration record which is kept with the built object (derived object). The configuration records are used by future builds to determine if the object is out of date and needs to be rebuilt. ClearMake examines the versions of all the elements referenced in the configuration record to determine if a rebuild is necessary. It does not rely on the "date-time-modified" of the object. The configuration records can be examined by the developer and used as a source of history information regarding any derived object.

Summary

During the course of the master memory controller chip design the RW_IO block went through many changes. The initial specification required that the block be long and narrow and cover the entire side of the chip that interfaces with the processor bus (see Fig. 6).

Later in the project, area became critical, and it was necessary to reduce the size of the block. An additional requirement on the master memory controller chip was that it had to support little endian byte order (least-significant byte is byte 0). The structure of the RW_IO block was ideally suited to implement the little endian functionality which required byte swapping the data returned from memory before delivering it to the processor bus (see Fig. 7).

The original specifications called for the main power and ground buses to pass through the center of the RW_IO block. Toward the end of the chip development the power buses changed and the V_{DD} bus had to be moved. This again required reconstructing the RW_IO block (see Fig. 8).

The use of standard cells and automatic place and route tools for the construction of the RW_IO block proved to be very advantageous. Because of the speed and flexibility of the tools the design team was not hindered when making changes to the high-speed interface block as they might have been if a full custom approach had been pursued. The most complex change to the block, which required rebuilding all of the bit slices and resizing and reconstructing the block, took only three days to complete, at which time the block was fully verified.

The design approach of using standard cells and semicustom layout techniques was favored over a data-path or full custom implementation because of its reduced risk. However, GDT schematics were generated for documentation purposes as well as a fallback position if a data-path solution had been required. The combination of standard cells, custom Cell3 layout, and accurate static timing analysis proved to be an effective solution for the RW_IO high-speed interface block.

Reference

1. S. Clayton, et al, "The Automation of Standard Cell Block Design for High-Performance Structured Custom Integrated Circuits," *Proceedings of the 1993 HP Design Technology Conference.*