

The PPC CD ROM Set : A Work in Progress

Jake Schwartz

Three Concepts Converge

Back in the early part of the year, Eric Rechlin announced the availability of his HP48 Archive collection on CD-R disk. This was the complete contents of his website located at <http://hp48.ml.org>, consisting of 75 megabytes of public-domain software and articles on the HP48. At only fifteen dollars, I couldn't pass up the chance at having instant access to all the goodies that had been developed by the HP48 user community over the previous eight years.

Roughly at the same time, I had been experimenting with encoding some music from ordinary CD disks into the recently-popular MPEG-1 Layer 3 ("MP3") format, which compresses the data at a rate of approximately twelve to one with very little if no loss in sound quality upon playback. I thought that at this rate, I could make my own CD-R disks and pack 12 to 15 albums on single disks, allowing me to enjoy lots of music playing on my laptop computer at my desk at work. This capability alone justified the purchase a recordable CD drive. So, I bought a drive and began happily making compressed music disks from my home CD collection.

Also at this time, I was doing some HP41 software for a friend with whom I had worked since 1991. A question arose, related to using the HP41 Plotter Module (written for use with the HP7470 two-pen HP-IL plotter): 'Could the HP41 use this same module with the RS-232 version of the plotter connected to the IL loop via the IL-RS232 converter?' I did not have the answer, and began combing through the manuals as well as old HP calculator newsletters from the early 1980's seeking an answer. Lots of time was spent, but no definitive answer was found. However, some of the reference material indicated that the plotter module software would have to be augmented with additional HP-IL commands from the Extended IL Module. This would allow the calculator to be able to "see" the plotter as just another ordinary HP-IL device on the loop.

Suddenly in the middle of February, the proverbial light bulb lit up over my head – it seemed like a good idea to make a CD of all the old HP calculator newsletters and offer it to the public at low cost. I was already in possession of all the old PPC Journals, CHHU Chronicles and HPX Exchange issues from 1974 through 1990, so an early study was done to count the pages and determine if all this could fit on a single 650-meg disk. Despite there being over five thousand pages, I still had my hopes up that some mechanism could be found to compress the data such that it all would fit. Also, given that a calculator conference was scheduled for late August, it seemed like a worthy goal to complete it by then so copies could be made available there. I ran the project idea past Richard Nelson via email and he was initially skeptical, feeling that this was too ambitious a task for the six months' time period. However, soon after my scanner and computer were put into action, his skepticism became enthusiasm.

Early Experiments

At the beginning, a format was sought for compressing the data from the newsletter pages in order to achieve a handful of goals:

- Data would compress small enough that all pages would fit on the single disk
- Expansion of the page data would be fast enough to permit navigation through the pages, issues, volumes and newsletters quickly via the computer's keyboard
- The quality of the pages needed to be such that they were not only viewable on the screen, but also readable in hard copy form

When various image-compression formats were attempted, some came close to the desired file size for an average page, but they did not look acceptable when re-expanded for viewing. The

format that came closest to achieving the aforementioned goals was JPEG, which compressed an "average" page of text down to around 180Kbytes. The printed copies of the pages didn't look really good, but this was the best I could find. So, the "official" scanning commenced.

Meanwhile, an environment was needed inside which the user could view the pages and also move from issue to issue. Originally, I had planned on emulating Eric Rechlin's efforts, whereby his HP48 Archive disk when loaded, would cause the user's web browser to start. (His entire disk is basically a complex web site etched in plastic.) This seemed logical, since HTML-based data would be able to be read by many different types of computers running under different operating systems. The idea was to have two HTML frames with the left one being a narrow vertical band of control buttons and the right one displaying the newsletter page images. Early on, it was realized that some relatively complex Javascript programming would be required to achieve the environment that had been envisioned. So, I began dividing my spare time between scanning the early 65 NOTES issues into the computer and learning the ins and outs of Javascript.

Adobe Acrobat to the Rescue

Many events happen simultaneously, and still another of those events was a running email conversation with Ron Johnson of the Chicago-area CHIP group. (They were formed in 1976 as the second satellite PPC Chapter following only the one in Dayton, Ohio.) Ron had been a fan of the Adobe Acrobat portable document (".pdf") file format for quite a while, and had been sending me articles and photos encoded as such. The freeware Acrobat Reader was on my own computer as well as the UNIX workstation that I used at my job, and it was always easy to pass info back and forth as well as through the email.

Since the .jpg files that were being generating from scanned newsletter pages were still not producing the ideal result (especially with substandard hard-copy), I decided to take the plunge and purchase Acrobat Exchange software for encoding the pages into .pdf format. This became a win-win experience: not only did the resulting page files print as clear as if they had been photocopied, but the files actually compressed *smaller* than their .jpg counterparts, averaging around 150 to 160K per page. I discarded the 175 pages that had already been scanned into .jpg format and started again with Acrobat.

One Disk Becomes Two

While scanned pages consisting entirely of text were compressing down rather nicely at a resolution of 300 dots per inch (which was the limit of my HP Scanjet 4p scanner), those with photographs had to be handled entirely differently. In order to capture the detail of photos, the scan mode had to be changed from one bit per pixel "black & white" to 4 bits per pixel grayscale. That shot the size of the page files up enormously, and it became apparent that many fewer pages would fit onto the disk. Rather than deliver less material than had been planned, it was decided to increase the project to a two-disk set. Upon letting Richard know of the situation that there would probably be a second disk that was partially full, he made a worthwhile suggestion - to attempt to fit the HP Key Notes issues and relevant handheld-related Hewlett-Packard Journal articles onto the available open space. Richard even had prior written permission to reproduce the HP Journal materials for user-club distribution. And so the ultimate contents of the PPC CD-ROM set as it exists today was finalized. It was March and up to this point only the first half-dozen volumes of the PPC Journal had been scanned.

Why Not Acrobat Capture?

Prior to this project, I had been scanning printed pages of text into optical character recognition (OCR) software and having relatively good outcome, with very few misread characters. In addition, Acrobat Exchange has a utility called "Capture", which performs OCR processing on scanned images prior to converting them to .pdf format. It was thought that the pages of the PPC Journal might be candidates for Capture, which would no-doubt reduce the page file size

significantly. Unfortunately, the Journal pages from the 1970's and 1980's consisted of printed copies of master sheets containing manually-typewritten articles pasted together. The variability of those typewriter ribbons between cloth and film, the different type styles and the light and dark reproduction caused test pages run through Capture to have only about half of their text recognized. Considering that manually inserting the missing letters would have taken up to an hour per page ruled out any sort of hopes for OCR on the Journal issues.

Keeping On Schedule

Originally, I had set out a goal for myself to complete the scanning of the *65 NOTES / PPC Journal / PPC Calculator Journal* issues by mid May, doing approximately two newsletter issues per day. This goal was missed by around two weeks, but still things seemed to be progressing far enough that all could be completed by conference time. Another goal was to have all the *Computer Journal*, *CHHU Chronicle* and *HPX Exchange* issues complete in time to present the material at the July 1st Chicago-area CHIP HP group meeting. With some intense weekends in June, this goal was achieved and the material was fairly well received. Through this time period, however, two changes to the navigation scheme were suggested that would impose some major revisions to the hyperlinks and bookmarks being inserted along the way. First, Richard Nelson suggested that links to issues should point to the first page of the actual newsletter and not the first page of the issue *file*, which usually consisted of a member letter or other supplementary material. This would enable faster searching for specific Journal items without having to repeatedly find page one of each successive issue. And second, the "jump pages", which were big tables of all issues where the user could click to go anywhere on each disk needed to be revised so the individual issue boxes would align horizontally with the month of publication. These two tasks were performed after returning from the Chicago trip.

Key Notes: The First Real Challenge

Following the major hurdle of finishing the 4700 pages of user-club newsletters, only the three Hewlett-Packard publications remained. With three-quarters of the pages scanned in at this point, it seemed as if it would all be down hill until *HP Key Notes* reared its ugly head. As anyone who has seen or read issues of *Key Notes* knows, the individual pages vary in background color from all white to gray to light blue. This provided a scanning nightmare, as each page required the scanner settings to be adjusted repeatedly until the text was clearly readable. Approximately three tries for each page was the norm for these 23 issues which average around ten pages apiece. One wonders if the variation in the page color was done intentionally to annoy people who might attempt to photocopy the issues.

Shoehorning the HP Journal

Following the insertion of *HP65 Key Note* and *HP Key Notes*, a little over 400 megabytes remained free on the second CD. After originally estimating that the number of pages required for the *Hewlett-Packard Journal* articles would be 300, it seemed a cinch to fit it all in. In fact, at this point it was considered that all the extra room could be used to include the eight HP Calculator Digests from the late 1970's to the early 1980's, whose pages are all printed in color. This turned out to be pi (π) in the sky, as the actual count of relevant HP Journal pages turned out to be over 600. Suddenly, I panicked that all the pages wouldn't fit on the disk. At the point when half the pages were scanned, only about 100 meg of available space remained.

Originally, the plan was to "pull out all the stops" with high-resolution scans of the *Journal*. Many of the pages and all of the issue covers were in full color, and I had hoped to be able to provide all of these pages in their original splendor. Unfortunately, a single page scanned in full color with high resolution required between 4 and 5 megabytes. There were 32 issues between 1968 and the present day with articles that deserved to be preserved. Just doing the issue covers alone would require almost 140 meg. (For consistency, it seemed reasonable to include the

covers of all included issues despite the fact that several covers were not related to the calculator stories within the pages.)

With 300 pages remaining, it was time to develop a strategy in order to significantly reduce the amount of space occupied by the first half of the pages. Luckily, there turned out to be a number of ways to save bytes. Issue covers that were not relevant to calculators were rescanned in black and white. Many of the pages that were only peripherally related to the subject material (such as tables of contents and pages of author biographies) were redone with lower resolution. Pages that originally seemed to require grayscale could be rescanned in one-bit-per-pixel black and white with the brightness adjusted to preserve detail. All these things shaved enough bytes to allow the remaining pages to fit on the second CD with barely a megabyte of space (out of 681 million) to spare.

It's Been Fun

Over the past month or so, the i's were dotted, the t's crossed, and many prototype disks were made. Repeated testing permitted several problems to be found and repaired. The six months' time was grueling but all in all, the outcome will hopefully prove to be a worthwhile payoff. Revisiting the pages of the newsletters of the past 25 years have been a rewarding experience for me and should provide as much insight as they do reference material for anyone else who uses the disks. Perhaps the articles, diagrams and photos will help us better appreciate the thousands of people who transported us from the slide rule to the present day.

What's Next?

Many "future projects" have been suggested, and these have been listed on the "PPC CD ROM" web page at <http://www.waterw.com/~jake-s/ppccdrom.htm>. Of course it all comes down to having enough spare time. However if we receive Hewlett-Packard's blessing I would like to follow the current disks with a set consisting of the user manuals for all the discontinued HP calculators from the HP35 to the HP48SX. Time and luck will tell.... Stay tuned.