SGP MEMO NO. 27

To: Pipeline Committee

From: R. T. Duquet

Date: April 7th, 1983

Subject: Timing

Recently one of our users tried to run a fairly large spectral line database through the pipeline system. There was some question about how long the map-making process took but the answer seemed to revolve around the time required to read the database from disk. Since we had not previously obtained timings for the DBFILL or DBUTIL programs using anything but small sample data sets, I decided to run a test to see how long things took in a real live situation. The following are the results.

The data was spectral line observations taken by user 443 (Butler Hine working with Arnold Rots) from noon to midnight on August the 8th of last year. The accumulated data consists of approximately 9 million channel-baseline values. Portions of the data were recorded on 3 different Modcomp tapes (V5224, V5229 and V5194). The observing regime was such that relatively long scans (370K values) of a source (NGC3031S) alternated with shorter scans (45K values) of two calibrators (0836+710 and 3C286).

DBFILL

All the data, for the source and for the calibrators, was filled into the pipeline database. Filling took place onto an empty disk pack. The scratch file used to communicate between the two phases of the filler process was placed on a different pack and even a different channel from the one used for the final database.

From start to finish, the time required to fill this dataset using the DBFILL program was 2 hours 23 minutes. This is a throughput rate of just slightly more than 1000 channel-baseline values per second.

The rate at which data was processed depended fairly strongly upon the number of values in a scan. For long scans the rate was roughly 75% faster than for short scans (1600 values/sec for source scans vs 900 values/sec for calibrator scans).

For spectral line data the DBFILL program notes the time at which the second sort phase begins and ends so the throughput of the second phase could be determined separately. (This is of interest because a version of the second phase sort program may be used for the on-line filler.) For large scans (370K values) the rate was 3800 values/sec. This dropped relatively slowly to 3200 values/sec for moderate size scans (132K values). For small scans (45K values) the rate was only 1500 values/sec. Unfortunately there were no scans containing between 130K and 45K values so it was not possible to determine at what size the rapid drop in throughput begins.

DBUTIL (Backup & Restore)

After the data set had been filled into a pipeline disk, I ran the DBUTIL program to back it up onto another tape. This was to answer the question as to whether or not is would be worthwhile to file two forms of the data, one on the standard Modcomp archive tape and the other a backed-up pipeline pigeonhole file. It was anticipated that restoring the pigeonhole file would be far more efficient than re-filling.

Since the database consisted of 32-channel spectral line values, the data for the source was distributed into 4 separate VIS files. Each VIS file contained the data for a different set of 8 channels but otherwise the files were identical. Timing figures were obtained by backing, verifying and restoring only one of the four identical files.

Each of the four files for the source contained 237,262 records or roughly 1.9 million channel-baseline values. All of the data fell within the first 40 (out of 256) pigeonholes. Approximately half of the data fell within the first 5 pigeonholes; 75% fell within the first 11. Each of these initial pigeonholes was fractured into roughly 40 non-contiguous blocks.

Several attempts to do the backup at 6250bpi failed because the data could not be read back to be verified (the read failure occuring at different places each time but invariably crashing the program!). As it turned out, the time required to backup one of the pipeline files described above did not differ noticeably when the tape was set to 1600bpi than when it was set to 6250bpi. In each case it took between 14 and 15 minutes to write the file and another 14 or 15 minutes to verify it. For the combined operations the throughput was therefore 1100 channel-baseline values/second. Restoring the backed-up file required 14 minutes (2250 values/sec).

In brief, it takes only half as long to restore a backed-up pipeline file of spectral line values as it would to refill it ab initio. For continuum data, which can be filled at twice the rate of spectral data, the filling rate and the restoring rate should be about equal.

Note however that the utility program DBUTIL, which is used for backup and restore, operates within 65K words of memory whereas the filler program, DBFILL, uses nearly 250K words of memory.