INTERNAL COMMUNICATION

Comments and Additions to Draft Document "Control Systems and Data Analysis for Large Gravitational Detectors" by A. Jeffries, P. Linsay and J. Livas.

M. E. Zucker

7/20/88

A. Scope of this memo

The draft prepared by AJ, PL and JL presents a useful outline for a program leading to a design specification, but is of course far from complete. It appears that the authors feel (and I concur) that at this point no very serious control system or data handling problems exist beyond the state of current available technology as we know it. Since digital technology advances very rapidly, it would be imprudent to embark now on a full-blown design of some particular data system concept; however, it should be the goal of this document to identify not only areas of difficult technology but also potentially costly subsystems. After commenting on the rather general remarks in the first few pages of the draft I'd like to add some thoughts on data analysis issues which may affect storage requirements and costs.

B. General Remarks

(Section 1) I think the question of specific hardware should be looked at, especially the use of parallel computing for certain searches. It might be nice to include a discussion of these topics as an appendix to this report.

(Section 2) I feel the "Guiding Philosophies" outlined in Section 2 are basically sound. I would like to add a few points to the list:

-- User Friendly Interface. A minimum of training should be needed for scientists and staff to be able to comfortably access information, inquire status, or employ the hardware and software provided. This will save time, allow "outsiders" and newcomers to contribute effectively, and inspire the experimenters' confidence in displayed results, an important factor in minimizing direct contact with receivers. This user confidence probably cannot be achieved without a continuing debugging and updating program, currently envisioned as a full--time task for a computer analyst at each site.

-- Expandability is an obvious plus, but I feel it is necessary to determine the potential scope of expansion; since many elements of the data system would presumably be shared among the various receivers, the hardware and software should be provided initially with capacity for the full complement of instruments (e.g., three full and three half-length receivers) envisioned at each facility, plus added capacity for future receivers (of greater sophistication) and redundancy for failure recovery.

-- "Laboratory Computer" support. Since the operations scenario envisions at least some receiver installation/engineering ongoing at any given time, sufficient computer support must be provided for testing and analysis of instruments and servos, debugging equipment, vibration measurements, optics analysis and testing, etc., as well as CAD capability for electronic and mechanical design and field revisions. These functions (with the possible exclusion of CAD work) may generally need to be integrated with the "observing" data system.

(Section 3) The hardware issues should be fairly easy to answer if given some better idea of boundary conditions. Sheri Smith has written up some ideas on "bare minimum" data storage and bandwidth requirements for the observations, but I think these were arrived at with a deliberate intent to minimize the quantity of stored data to cut a perceived big cost. I'm still unsure about that perception, and would be in favor of more exploration of the storage costs (e.g., VCR tapes). I haven't prepared anything yet on this. I'd prefer to reorganize this section with special attention to logical divisions in the hardware functions; for example, the timing system and vacuum system monitoring will presumably be laboratory-wide and common to all experiments, while digital servo systems might be receiver-specific. A further point; the possible items in subsection 5. are really too numerous and detailed to enumerate effectively at this stage, and in fact are subject to radical change. I'd suggest holding off on refining this list until we have a better idea of storage limits. The list in subsection 6. is again mixed among fairly receiver-specific items and items which affect all observatory functions; the control of vacuum valves needs to be considered in a different context, since it may require site-wide coordination.

The section on hardware issues needs more structure; a more streamlined plan of attack for the design specification is required, and I would like to coordinate this with ongoing work on "receiver/facility interfaces." Some specifics regarding transmission of receiver signals along the arms will be presented in an upcoming memo on the interface question.

(Section 4) As I mentioned above, I'd like to elevate user-friendly software to the status of a guiding philosophy. To the list of software issues presented I'd like to add a couple of useful functions;

Item 5. We should build and maintain of a "library" of data analysis tools, developed for searches not routinely done on-line but readily applicable to archived data (or the fresh stream) whenever something comes up (a supernova, a new fast pulsar, an X-ray burst, etc.) with absolutely zero programming and debugging. In this way we can be ready to instantly perform those searches beyond our on-line capability whenever special conditions dictate.

Item 6. We should arrange for duplicate hardware so that computer analysts and scientists can develop and test data analysis routines concurrently with continous on-line observing tasks, and add new or improved on-line experiments as they become available. I'm pretty sure that software development will need to progress in stages, as will the receiver installation, and simple bare-minimum algorithms employed at startup will eventually be replaced by more and more sophisticated codes (the same phased installation will probably be appropriate for any special-purpose hardware developed).

(Section 5) Data analysis requirements are well discussed insofar as they affect data rates, mass storage, timing, and computing requirements. I would like to open discussion on some of the points, although the guidelines in the draft are quite reasonable.

For impulsive and coalescing compact binary sources, the conclusion that current (and reasonably priced) technology can do the job appears well founded; I don't foresee excessively costly or unwieldy hardware for these on-line tasks, even for what I would consider "luxurious" coverage of most expected source waveforms with near optimal signal-to-noise.

Massimo Tinto and Yekta Gursel have developed a clear formalism and efficient numerical implementation for the inverse problem (assuming three equivalent detectors at separated locations). It appears at this stage that the computations involved may be rather intensive; development is proceeding with the assumption that candidate events are identified by standard (e.g. matched filter) techniques and subsequently examined to determine the source position and parameters. Preliminary tests indicate that in this mode many candidate events per day could be examined with a moderately fast computer like a Ridge or Sun 4/260. The procedure is highly conducive to parallel processing (splitting the sky up into regions, for example) so even modest parallel machines could greatly enhance the speed. Massimo and Yekta should have some results soon on periodic and stochastic sources as well.

C. Additional Comments

In the subsection on periodic source searches, it is estimated that one needs to oversample by a factor of ten to avoid "spurious modulation" in interpolating the Doppler-corrected time series. I would like to see more of the rationale behind this and the behind the additional recommendation that Doppler-correction be done by special purpose hardware; I'm not sure on either of these points. For one thing, Yekta and Bob Spero assure me that cubic spline interpolation requires only about 10 operations per point; surely this overhead is negligible compared, for example, to the burden of reading data off slow mass storage. Because the data is band-limited to the Nyquist interval, adjacent samples are unlikely to differ by much; as a result, linear interpolation might provide sufficient accuracy. Even no interpolation at all (just substituting at each Doppler-corrected sample time the nearest uncorrected value) would only degrade the SNR a small amount. After all, this is just equivalent to a phase modulation of the source with a maximum deviation of half a sample period. Even at the highest accessible search frequency, a modulation like this should rob less than half the power from the carrier peak.

Furthermore, if there really is a problem, we may have reason to look into Doppler correction in the frequency domain; convolving a complex filter equal to the transform of the the Doppler shift (a time-dependent complex phase factor) with the raw data transform. This could have additional advantages, as the signal power is spread among only a sparse set of sidebands due to the periodicity of the Earth's motions. It could also be economically combined with peak detection. In any case, it would obviate the need for time-domain Doppler correction. This concept demands further analysis.

The reason for stressing these points is that oversampling increases data storage and handling tasks by up to an order of magnitude: even if the above remarks are not sound, I'd like to stimulate thought toward avoiding this burden (which is, after all, only assumed for computational convenience) if at all possible.

Finally, I feel the need for special-purpose hardware has to be better established in this case: big array processors and parallel machines may soon be commercially available, and (in keeping with our Guiding Philosophy) I'd prefer to consider homebrew gadgets only as a last resort.