AN ENSEMBLE PULSAR TIME

Gérard Petit and Claudine Thomas
Bureau International des Poids et Mesures
Pavillon de Breteuil
92312 Sèvres CEDEX, France
and
Patrizia Tavella
Istituto Elettrotecnico Nazionale G. Ferraris
Strada delle Cacce 91
10135 Torino, Italy

Abstract

Millisecond pulsars are galactic objects that exhibit a very stable spinning period. Several tens of these celestial clocks have now been discovered, which opens the possibility that an average time scale may be deduced through a long-term stability algorithm. Such an ensemble average makes it possible to reduce the level of the instabilities originating from the pulsars or from other sources of noise, which are unknown but independent.

We present the basis for such an algorithm and apply it to real pulsar data. It is shown that pulsar time could shortly become more stable than the present atomic time, for averaging times of a few years. Pulsar time can also be used as a flywheel to maintain the accuracy of atomic time in case of temporary failure of the primary standards, or to transfer the improved accuracy of future standards back to the present.

1. MILLISECOND PULSARS

Pulsars are neutron stars which spin rapidly and have a strong magnetic field. This generates a directional beam of electromagnetic radiation, which makes the pulsars detectable when the Earth happens to lie on the path of the beam. In 1982 the first element of a new class of pulsars, millisecond pulsars, was discovered, PSR1937+21 [1]. These millisecond pulsars not only spin much more rapidly, but have distinct physical properties that make them essential to an understanding of the evolution of pulsars.

It was quickly recognized that the period of rotation of PSR1937+21 is very stable. Since 1982, astronomers have carried out timing observations which measure the time of arrival (TOA) of the radio pulses relative to an atomic time scale. They have also discovered many more millisecond pulsars, which are now regularly observed at a number of observatories. The technique allows measurement of the TOAs with a precision of order 1 μs. Such a precision maintained over the whole period of observation (many years) makes it possible to reach a level of stability unequalled by any other technique. This opens new paths in astronomy and astrophysics: detection of gravitational waves, probe of relativity in strong fields and high velocities, detection of planets orbiting pulsars, etc [2].

As the measurements are referred to atomic time (AT), pulsar analysis demands that this reference time scale be as stable as possible. Conversely it must be possible to transfer the stability of the rotation of pulsars to atomic time if the former appears to be better for certain averaging times.

2. PULSAR TIME

2.1. Astrometric analysis of pulsar data

The measurement system detects the TOA of pulses at the phase center of a radiotelescope. The TOAs are labelled in a local atomic time scale, and can later be referred to any time scale by the appropriate time transfer technique. It is possible to relate the TOAs to physical parameters like the position of the pulsar, that of the Earth, the spin rate of the pulsar etc. The goal of pulsar data analysis is to determine the set of physical parameters that best fits the observations.

To give a simple representation of the physical relation, we can consider the ideal case of a pulsar with a constant period, without motion relative to the solar system, and fictitiously observed at the geocentre. In this case the TOA observations would be regularly spaced except for an annual variation due to the orbital motion of the Earth. This one–year signature would determine the astrometric position of the pulsar, and the period would be determined as the average spacing between TOAs. The true situation is of course more complicated [3], but the general outline is similar.

2.2. Pulsar Time

The astrometric analysis provides a deterministic model to predict the arrival times of pulses. This is equivalent to defining an individual time scale PTi for pulsar i as follows: if A is the event "arrival of pulse n of pulsar i at a given observatory", $PTi(A)$ is the TOA computed from the model for this pulse. This way we define a time scale which is usable, although a bit cumbersome: with a radiotelescope and an adequate measurement system, any event can be compared to the arrival of a pulse, at least in theory, just as any event can be compared to the 1pps of an atomic clock.

For a given pulse, $PTi(A)$ is the predicted TOA and AT(A) is the observed TOA referred to the atomic time scale AT. The residuals of the timing data analysis, computed as observed minus predicted values, represent the difference of the two time scales AT minus PTi at the dates of the observations.

2.3. Relationship between AT and PT

In the present state of pulsar observations, the measurement uncertainty is as low as 0.2 μs in the best case [4], but more generally is at the level of 1 μs. It is clear that the instabilities of the present realizations of AT can be larger than 1 μs for averaging times of a few years. Indeed the fractional frequency instability of the present best realization of an uniform atomic time is estimated at about 2×10^{-14} [5] and statistical analysis of the timing residuals of PSR1937+21 over several years of observation [2] indicates that the instability of AT–PT reaches similar values for averaging times larger than one year.

It is thus tempting to assume that the intrinsic stability of PT (ie. the stability of the rotation of the pulsar) is better than that of AT, and to use PT as a reference. In doing this, three problems arise.

First, it is not possible to define PT independently from AT. This is due to the variation of the rotation period of the pulsar, which cannot be described by a physical model. The variation of the period is assumed to be linear with time, but its rate can be determined only by reference to another time scale, that is to AT [6].

Second, AT–PT is the list of residuals to a fit which determines the physical parameters defining PT. All but three of the fitted parameters correspond to short term periods (maximum one year), and the remaining three correspond to a second–order polynomial of time. Thus such a fit filters from the residuals all terms with short periods (up to one year) and long periods (of order the time span T), so the instability of AT–PT can validly be estimated only for averaging times between about 2 years and T/2. From this, we can infer the instability of AT only for averaging times long enough to bring the measurement noise down to an instability level comparable with that of AT. These times may range from six months to several years depending on the measurement system and the pulsar.

Finally, several other sources of noise affect AT–PT, in addition to the intrinsic noise of AT and of the rotation of pulsars. Among them are interstellar propagation effects, uncertainties in the ephemeris of the solar system, and gravitational waves. All these contributions have identifiable signatures so that in theory it is possible to discriminate among them, as proposed with the concept of the pulsar timing array [7]. But for our purpose of finding a stable time scale with which to compare AT, it is much more convenient to assume that all sources of noise except AT are independent for different pulsars, and to average them by defining PT as a weighted average of the individual scales from several pulsars PTi. To do this we have to devise an appropriate stability algorithm, a procedure similar to that used for atomic clocks.

3. ALGORITHM FOR AN ENSEMBLE PULSAR TIME SCALE

The leading idea is to define a suitable algorithm that allows the computation of an ensemble pulsar time PT with the best long–term stability (integration times from 2 years up to T/2), as discussed in section 2. In pursuit of this aim, the following choices seem appropriate: PT is computed *a posteriori* with the largest possible data sets;

- PT is a weighted average of each pulsar contribution;
- weights are chosen according to the long–term stability, over some years, of each pulsar;
- precautions are taken to avoid the injection of unwanted noises by residual deterministic trends.

An algorithm corresponding to these characteristics has been devised and tested on simulated data, as reported in [8], and it is here adopted for the analysis of real observation data.

3.1. Preparation of data

As seen in section 2, the pulsar observation data are the results of an analysis in which the different parameters of a model have been estimated. To the aim of constructing an ensemble pulsar time, further analysis is necessary. It is desirable that the raw timing observations are made available and reduced with the same theoretical model to ensure the consistency of residuals. When the only available data are the residuals, it is desirable that the parameters of the models are provided to check that no important differences exist between the models used. In particular, if the residuals concern the same pulsar as observed by different observatories, the differential drift of one data set versus another, if any, has to be removed.

The residuals obtained are (AT–PTi) where AT is a chosen reference atomic time. They should contains only random components, apart from the residual deterministic trends due to uncertainties in the estimation of the parameters of the model.

In the computation of a time scale, equispaced data are convenient. Since we are interested in the long–term behavior of pulsar time, it is possible to construct a set of equispaced data with a suitable short–term average of the available residuals. We have chosen to use residual series with data spaced by 0.1 year. In order to prepare such series we process the available residuals with a moving average on a 0.2 year interval (containing at least three data points) centered on the date of interest. After such processing the resulting equispaced data are not really the residuals of observations, but an average of them, and their stability for integration times lower than 0.2 year is biased. Since we are interested in the stability for integration times larger than one year, this is not a restriction.

3.2. Ensemble algorithm

In analogy with what is done for atomic time scales, the definition of the ensemble pulsar time PT is the weighted average of the basic measurements (AT–PTi). It is defined in the form of the difference AT–PT, as

$$AT - PT = \sum_{i}^{8} w_i (AT - PT_i)$$

where w_i is the relative weight assigned to pulsar i. Weights are defined as inversely proportional to the instability of each pulsar for an integration time of a few years. This can be realized by taking for example the inverse of the Allan variance for 2.5 years, normalized to unity.

Since PT is computed *a posteriori*, when the complete set of data is available, each pulsar enters the ensemble with a fixed weight corresponding to the estimated level of instability.

To optimize the long–term stability, it appears important to have a good estimation of the drift of the pulsar period in order to avoid a residual drift occulting the long–term stability. This requires an observation time long enough to smooth the measurement noise (white phase noise) and to reach the long–term stability floor of the atomic reference time scale. At present, using TAI as reference atomic time, this level is estimated to be 2×10^{-14}. If the instability of TAI–PTi reaches this level, the residual drift will not degrade the stability for the integrating time of interest. If for any reason other sources of noise exceed this level and the long–term instability of TAI–PTi does not reach the above floor, the long–term behavior of pulsar i is treated as unstable and this pulsar enters the ensemble with a lower weight.

In the present test the weights have been estimated by computing the Allan variance of TAI–PTi. If more data sets were available it might be worth estimating the instability by an N–cornered hat technique or by reference to an equal weight average, as discussed in [8].

The weight of each pulsar is fixed, but the number of observed pulsars can change as new pulsars are discovered or an observatory interrupts the pulsar timing. When the number or weight of clocks changes in an atomic time scale, suitable corrections are added[9] to avoid time or frequency jumps. In case of an average of pulsar residuals, all the deterministic trends have already been removed, and the only possible effect is a time jump resulting from computations with ensembles having a different number of residuals. So the removal or entry of a new pulsar is accompanied by a time correction that ensures the continuity of the pulsar ensemble scale. Such correction, a, is defined as

$$AT - PT = AT - (PT' + a),$$

where PT' is the new ensemble pulsar time computed with the data available after the change. It is easy to see that the correction, a, is just the difference in time between the new $PT\prime$ and the old one.

4. APPLICATION TO REAL DATA

4.1. Available data

In this first experiment with real pulsar observations, the available data are the published residuals of the timing measurements performed at Arecibo Observatory [10]. They concern PSR1937+21 for the period 1984.9 to 1991.2, PSR1855+09 for the period 1986.2 to 1991.2 and PSR1957+20 for the period 1988.4 to 1991.2.

The residuals have been processed according to the procedure described above to obtain equispaced data with an interval of 0.1 year. Since all data come from the same observatory and have been analysed with similar models, the residuals have been used without further processing.

The residuals used TAI–PTi are reported in Figure 1 and their estimated Allan deviations in Figure 2. From the instability behavior it can be seen that pulsars 1937+21 and 1855+09 reach comparable instability levels for an integration time of about 2 years, while 1957+20 is ten times worse. For this reason the weights adopted for 1937+21 and 1855+09 are equal, while the weight of 1957+20 is 100 times lower (For the period in which all three pulsars were measured, the weights are respectively .497, .497, .006)

4.2. Results

The ensemble pulsar time PT, computed with the real data described above and compared with TAI, is shown in Figure 3. From the instability analysis (Figure 4) it can be verified that the ensemble PT is more stable than any single PTi for integration times in the range 1 to 2.5 years. In this region the Allan variance of TAI–PT is almost fixed at the level 2×10^{-14}. Assuming that, for the integration time of interest, the instability of TAI–PT is mostly due to that of TAI itself, it is interesting to examine the long–term behavior of TAI–PT. To do that, a Vondrak smoothing has been applied to smooth the noise due to Fourier components with periods lower than 1.5 year.

The result may be seen in Figure 3. It shows a cubic signature which is mostly due to the fact that a second order polynomial has been removed, nevertheless some information about the behavior of TAI can also be inferred. This is discussed in section 5.

5. DISCUSSION

5.1. Estimation of the stability of TAI and PT

For averaging times from 1 to 2.5 years, the measured fractional frequency instability of TAI–PT is about 2×10^{-14} (Figure 4). As this is also the estimated long term instability (and inaccuracy) of TAI **[11]**, we have a good indication that PT and TAI can both be assigned an upper limit of 2×10^{-14} for 2–year stability. We have tried to confirm this indication by performing two simple tests on the present data set.

First we consider how the uncertainty in the determination of the time derivative of the pulsar period $\dot{P}$ (linked to a quadratic term of time in the residuals) can influence the estimated stability of each pulsar time, and consequently its attributed weight and the average PT. From the 1937+21 data, the uncertainty on $\dot{P}$ is estimated to 2×10^{-25} s/s **[10]**. This corresponds to a quadratic term of 6×10^{-23} s^{-1}, which has a theoretical 2–year Allan deviation of 0.7×10^{-14}. If we arbitrarily add such a quadratic term to the residuals, we observe that the 2–year frequency instability of this pulsar is not significantly changed, and the instability of the average is very similar.

Second we try to refer the pulsar data, and PT, to another atomic time scale. In the computation of TAI, the BIPM first generates a free–running time scale named EAL. TAI is then derived from EAL by steering its frequency to that of the primary frequency standards. As a test we have used EAL as a reference for pulsar data, and computed EAL–PT. It is interesting to note that the comparison with TAI–PT provides two hints which favor TAI, as might be expected. The difference between TAI and EAL over the period of reference arises mainly from a number of frequency steerings, after mid–89, that bend TAI "upwards" (to a net amount of 1 μs at the end of 1991). It can be seen in Figure 5 that this bending makes TAI–PT wander less than EAL–PT, a kind of visual indication that the steerings acted in the right direction. This is confirmed by the 2–year Allan deviation which is 10% higher for EAL–PT. Although this is not statistically very significant, it may provide the first evidence that the steering of TAI resulted in a time scale more stable in the long term, because it is more accurate.

This application to real data has limited value because only 3 pulsars were used, and for only 2 of them has TAI–PT reached a level of stability comparable with that of TAI. Furthermore the dataset with 2 pulsars or more covers a period of only 5 years. Nevertheless the outcome of this work looks promising because we have reached the level where we can infer some information about TAI, and the situation is one which will evolve rapidly. From the present programme of pulsar observations we anticipate that at the end of the century we shall have 10 to 16 years of observations on 6–8 pulsars, half of them yielding data comparable to PSR1937+21, ie. showing a 2–3 year Allan deviation of a few parts in 10^{14}, the remaining ones reaching this level after 5–10 years. This will considerably improve the stability of the ensemble average.

5.2. Transfer of the accuracy of atomic time

It is possible to take advantage of the long term stability of PT to transfer the accuracy of atomic time from one period of time to another, provided continuity in the pulsar observations is maintained. This is because, under the proposed scheme, the random part of the long term instability of PT can be decreased by averaging to a level much lower than that of the uncertainty in the present realization of the atomic second.

One application could be to use PT as a flywheel to maintain the accuracy of atomic time in the event of a temporary failure of the primary frequency standards. Similarly, when a future frequency standard with improved stability has been in continuous operation for years (and eventually provides a new definition of the second), PT will allow a backwards transfer, making it possible to evaluate the accuracy of our present atomic time scales. In this case, however, a limitation could arise from the random part of the instability of PT itself, which may be worse than that of the new standard for long averaging times. Such a situation is illustrated in Figure 6 where we show the result of a simulation. We have generated an atomic time scale which has an accuracy in the 10^{-14} range for 15 years and in the 10^{-16} range for the next 15 years, and a pulsar time PT which has random errors in the 10^{-15} range. When PT is referred to AT over the whole period, P and $\dot{P}$ can be determined so well that, over the first 15 years, AT–PT reveals the inaccuracies of AT (dashed line). In contrast when only the first 15 years are available, AT–PT only reveals the instabilities in AT (dotted line).

6. CONCLUSION

A millisecond pulsar can provide a time scale whose long term stability could be comparable with, or even better than, that of the present atomic time. Using data from many pulsars, it is possible to derive an average pulsar time scale that has a stability better than atomic time and better than the time derived from individual pulsar data. This improvement holds for averaging times from above one year up to about half the period of observation. A simple algorithm to realize this goal has been described.

A tentative application of the above procedure to real data yields limited results because very few pulsars have been observed, and the time span of the observations is only a few years. Nevertheless it seems that such a realization of pulsar time reaches the level of instability of atomic time. Current programme of observations make it possible to have enough data at the end of the century to estimate the instabilities of the present atomic time. If a more accurate atomic time scale is then available, it will also be possible to determine the present inaccuracies of atomic time.

REFERENCES

[1] D.C. Backer, S.R. Kulkarni, C. Heiles, M.M. Davis, W.M. Goss, *"A millisecond pulsar"*, Nature, 300, 615–618, 1982.

[2] J.H. Taylor, *"Millisecond pulsars: Nature's most stable clocks"*, Proc. IEEE, 79, 7, 1054–1062, 1991.

[3] R. Blandford, R. Narayan, R.W. Romani, *"Arrival-time analysis for a millisecond pulsar"*,

J. Astrophys. Astr., 5, 369–388, 1984.

[4] J.H. Taylor, *"Pulsar timing and relativistic gravity"*, Phil. Trans. Roy. Soc., 341, 117–134, 1992.

[5] A. Bauch, K. Dorenwendt, B. Fischer, T. Heindorff, E.K. Müller, R. Schröder, *"CS2: The PTB's new primary clock"*, IEEE Trans., IM–36, 613–616, 1987.

[6] B. Guinot, G. Petit, *"Atomic time and the rotation of pulsars"*, Astron. Astrophys., 248, 292–296, 1991.

[7] R.S. Foster, D.C. Backer, *"Constructing a pulsar timing array"*, Ap. J., 361, 300–308, 1990.

[8] G. Petit, P. Tavella, C. Thomas, *"How can millisecond pulsars improve the long term stability of atomic time scales"*, Proc. 6th European Frequency and Time Forum, 57–60, 1992.

[9] B. Guinot, *"Some properties of algorithms for atomic time scales"*, Metrologia, 24, 185–198, 1987.

[10] M.F. Ryba, *"High precision timing of millisecond pulsars"*, PhD dissertation, Princeton University, June 1991.

[11] Bureau International des Poids et Mesures, *"Annual report of the BIPM Time section"*, Vol. 4, 1991.

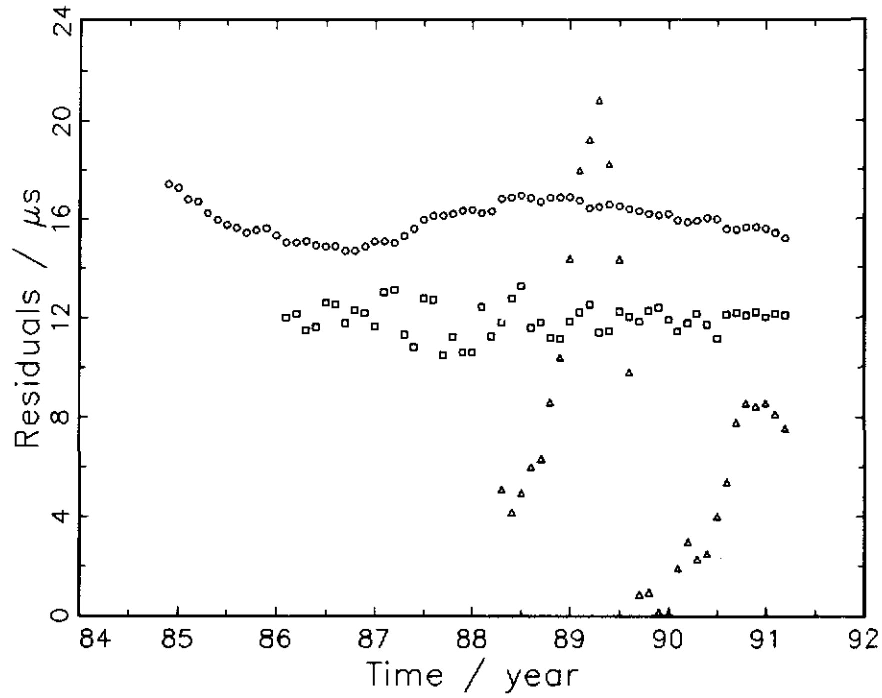

FIGURE 1: Timing residuals referred to TAI for 3 pulsars: 1937+21 (circles), 1855+09 (squares) and 1957+20 (triangles). The residuals have been averaged to equally-spaced data. For clarity, an arbitrary constant has been added to each data set.

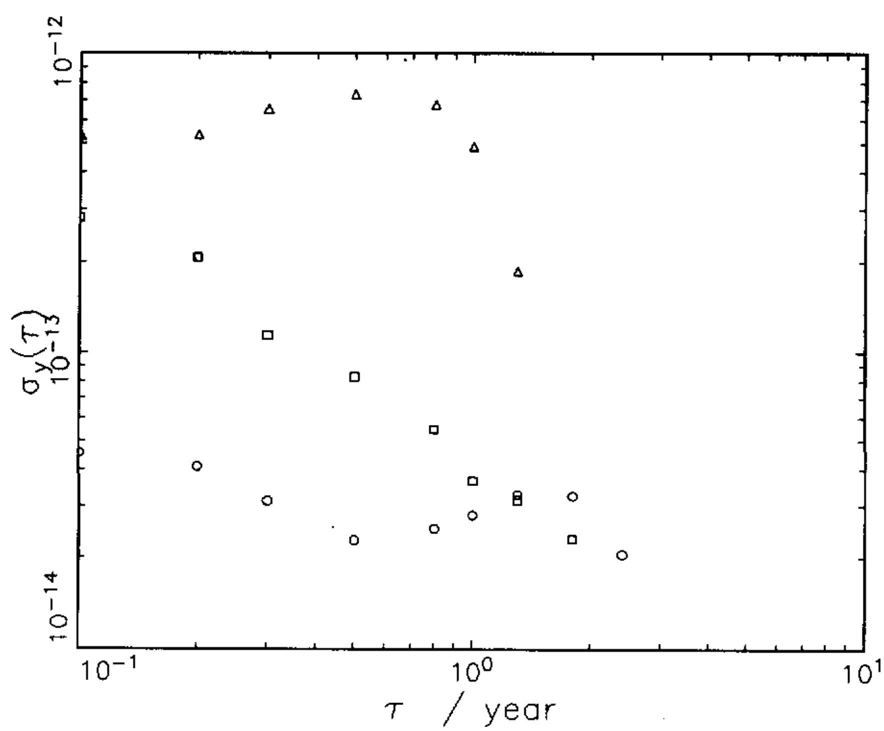

FIGURE 2: Square root of the two-sample Allan variance $\sigma_y(\tau)$ for the pulsar data of Figure 1 (The same symbols have been used).

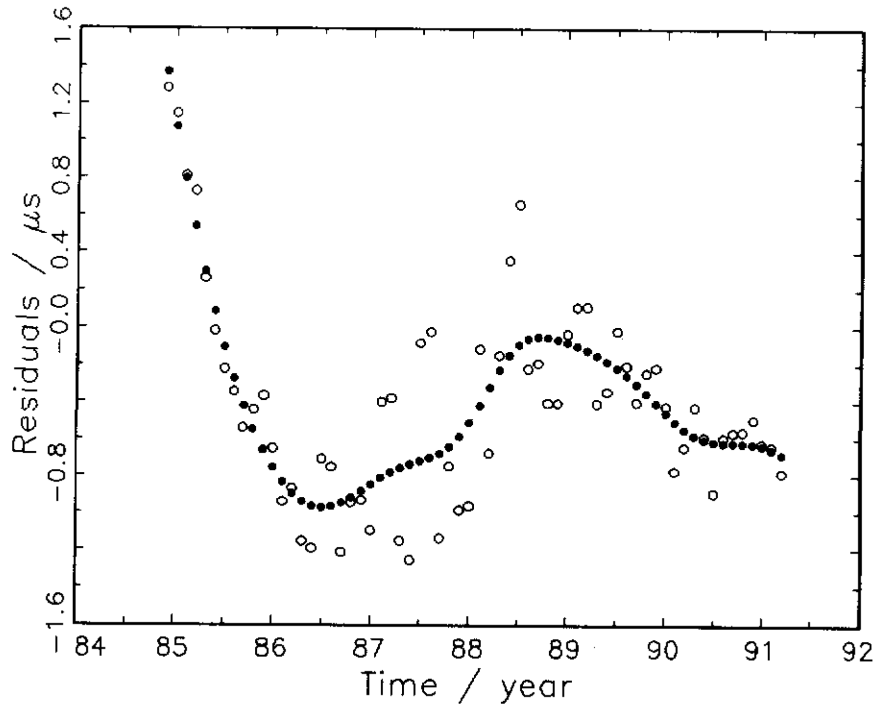

FIGURE 3: Difference TAI - PT, where PT is the ensemble pulsar time. Filled circles correspond to a Vondrak smoothing with a cut-off period of 1.5 year applied to the data.

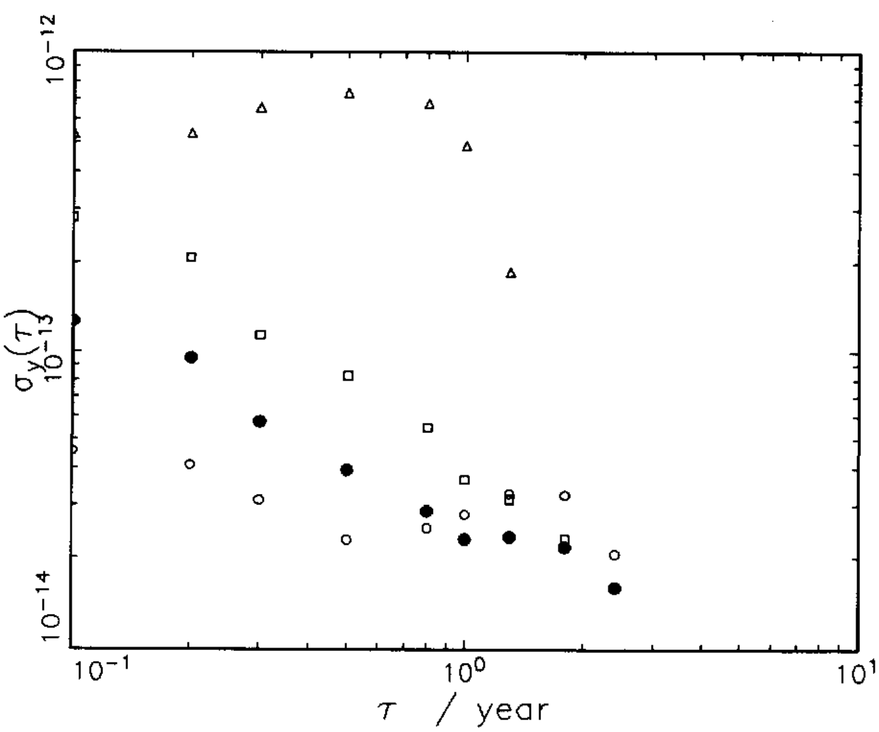

FIGURE 4: Square root of the two-sample Allan variance $\sigma_y(\tau)$ for TAI - PT (filled circles). Values for each pulsar from Figure 2 are also reported.

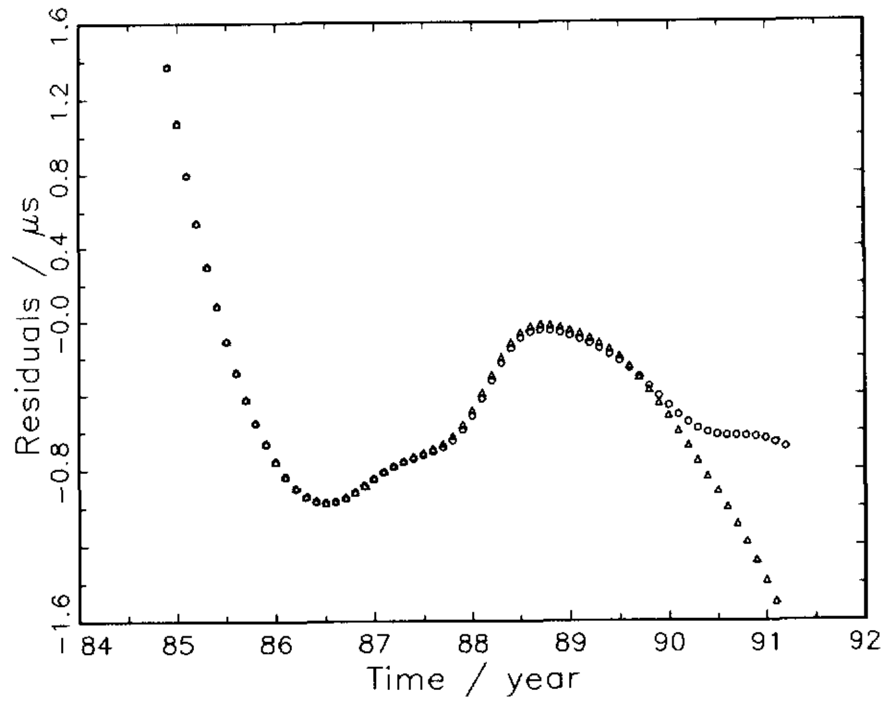

FIGURE 5: Differences TAI - PT (circles) and EAL - PT (triangles). A Vondrak smoothing with a cut-off period of 1.5 year has been applied to the data.

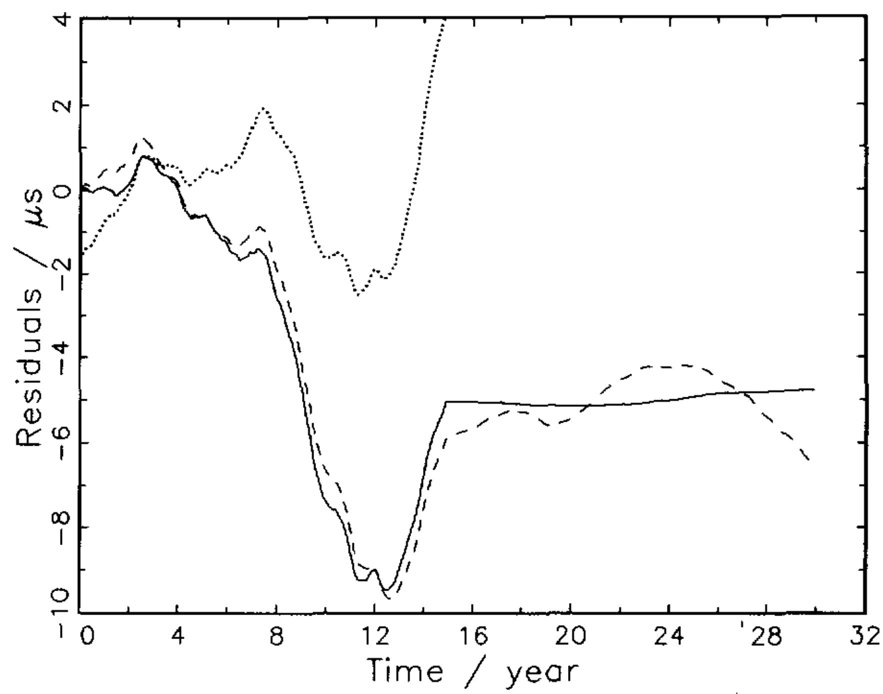

FIGURE 6: An atomic time AT with 10^{-14} accuracy for 15 years and 10^{-16} accuracy for 15 more years is simulated (solid line). A simulated pulsar time PT with 10^{-15} stability is referred to AT over the 30 years (dashed line). It allows to transfer the accuracy of AT from the last 15 years to the first 15 years to the 10^{-15} level. When PT is referred to AT over the first 15 years only (dotted line), it merely allows to estimate the instabilities of AT.

QUESTIONS AND ANSWERS

D. Matsakis, USNO: It seems like it doesn't meet the criterion of common sense, but here is my argument. You take the best pulsar and say that it's maybe 2 or 3 times worse than what you think the atomic clocks are. Then you average in another pulsar which you know is worse and you average it in with the same weight which might be the wrong way to average it and you say you're as good. That doesn't seem to be quite right.

G. Petit, BIPM: The ways we have chosen are representative of the stability of the pulsar for averaging times of say 2 years here.

D. Matsakis: I am not questioning the weight. I am just saying you will expect to maybe you do square root of two better or something like that; but now you're saying your direction here is good, and when you look at the two curves, the long term things, they don't really seem to agree with each other.

G. Petit: Which two curves?

D. Matsakis: No, the one you show with the three pulsars. The one that was very bad which seems to have a systematic problem to it, if I may comment there.

G. Petit: You mean the residuals themselves?

D. Matsakis: Yes. Well, all the stability on the first, where they showed all three pulsars together.

G. Petit: This one?

D. Matsakis: Yes, that one. It doesn't seem to me like those two pulsars, the two good pulsars agreed with each other more than they agreed with the atomic clocks. If I were to look at that I would say that the millisecond pulsar; the best pulsar is the deviant one of the three; the worst one, on the basis of the long term thing; long term residuals.

G. Petit: We cannot say the stability will obtain $2\times10^{-14} \pm 0.001$. What we can say is the level of stability that is achieved by using two pulsars which have comparable stability of three years. This is a level of stability which is achieved is comparable to that of TAI. So that pulsar time itself is not lost in TAI. So it is at least as good.

G. Winkler, USNO: Is it possible to apply the three cornered hat method to resolve these four differences?

G. Petit: Yes, it's a way to estimate stability of which pulsar given the difference of atomic time minus pulsar time. You can in fact with the cross differences with pulsar time, 1 minus pulsar time and estimate that each stability of each pulsar time is squared. I've mentioned that we have not used this for this case because the case that two pulsars are the same weight and either one was (?) but clearly for several, or if you have many pulsars, it is a way to do it.

D. Allan, Allan Time: Another question which I think is very relevant here. If you look at the history of atomic time keeping frequency standards have improved about an order of magnitude every seven years. If you project the year 2000, we should be another order of magnitude better. We know of devices that should be at least that good if not better than 10^{-15} to 10^{-16} by that time and if you now look at the fundamental problem with pulsar metrology, it is the measurement noise. Even with the new upgrade at Arecibo, which is costing ten million dollars, we anticipate the noise level to go down to 100 nanoseconds. If I do optimum statistical processing using modified $\sigma_y(\tau)$, I expect to see a clock projected into the future at that level of an integration time of about 200 years. I'm limited by the measurement noise in other words.

G. Petit: Yes, but if you consider that now it is expected to find one pulsar in every one hundred square degrees of the sky, you can anticipate to have several hundred pulsars to average. So pulsar time could gain a lot of stability just by statistical arranging.

D. Allan: So you're saying that by sheer numbers if I had 100 pulsars then I would have an improvement of the factor of 10 from statistical independence, is that what you're saying?

G. Petit: Yes.

D. Allan: Even with that if you drop that down to 20 years, the projected improvement in atomic clocks would out strip anything you can achieve from pulsars.

G. Petit: That is true but what I am also saying is that we can take advantage of the whole set of pulsar data and we can use the future improvement of atomic clocks to gain information at the present atomic time. That is also a use of millisecond pulsars.

D. Allan: I am happy to see the data. I am just projecting that by the year 2000, you may be left in the dust.

G. Petit: Well anyway, why do you assume I would be happy to have 10^{-15} time scales to get better data.

Question: Question about whether you've taken relativity and inertial frame effects into account for pulsar time?

G. Petit: That is taken into account in the last phase, which we have not developed because, with a simple model, it is actually much more complicated. That analyzes here to fit the model to the observation, takes into account everything which has to be taken into account; including relativity and whatever you want.