Colloquium: Comparison of astrophysical and terrestrial frequency standards

John G. Hartnett[*] and Andre N. Luiten

School of Physics, University of Western Australia, Crawley, Western Australia 6009, Australia

The stability of pulse arrival times from pulsars and white dwarfs have been reanalyzed using several analysis tools for measuring the noise characteristics of sampled time and frequency data. The best terrestrial artificial clocks are shown to substantially exceed the performance of astronomical sources as timekeepers in terms of accuracy (as defined by cesium primary frequency standards) and stability. This superiority in stability can be directly demonstrated over time periods up to 2 years, where there is high quality data for both. Beyond 2 years there is a deficiency of data for clock-to-clock comparisons, and both terrestrial and astronomical clocks show equal performance being equally limited by the quality of the reference time scales used to make the comparisons. Nonetheless, the detailed accuracy evaluations of modern terrestrial clocks imply that these new clocks are likely to have a stability better than any astronomical source up to comparison times of at least hundreds of years. This article is intended to provide a correct appreciation of the relative merits of natural and artificial clocks. The use of natural clocks as tests of physics under the most extreme conditions is entirely appropriate; however, the contention that these natural clocks, particularly white dwarfs, can compete as timekeepers against devices constructed by mankind is shown to be doubtful.

CONTENTS

I. INTRODUCTION

It has been asserted numerous times in the literature that pulsed astrophysical sources could be (Allan, 1987; Allan *et al.*, 1989; Hama *et al.*, 1989; Hotan *et al.*, 2006; Jacoby, 2005; Manchester, 2004; Matsakis *et al.*, 1997; Petit and Tavella, 1996; Rawley *et al.*, 1987; Roy *et al.*, 2005; Sazhin, 1989; Taylor, 1991; Verbiest *et al.*, 2009; Zhong and Yang, 2007) or are (Kepler *et al.*, 2005; Mukadam *et al.*, 2003) the best clocks[1] in the Universe. This belief is even more widely perpetuated in the popular science literature. In this article we contend that this assertion is not correct and, in fact, today, by all sensible measures of clock performance, artificial clocks have substantially better performance than natural astrophysical clocks, at least out to time scales of a few years. We believe that though there may have been some basis for these

claims in the past, this is no longer the case, and in some cases, the claims have been based on an erroneous analysis leading to an incorrect comparison of the relative stability of astronomical pulse arrival times and fluctuations in the frequency of artificial clocks.

In this paper we present formulas for making comparisons between artificial clocks, white dwarfs, and pulsars and also present new analysis of several millisecond pulsars together with an analysis of a set of astrophysical sources that have been represented as having the potential to function as high-performance clocks. However, it should be noted that in all cases where we refer to pulsar (or white dwarf) stability we are actually referring to the comparison of the astrophysical clock signal against a terrestrial reference clock and/or time scale developed from an ensemble of many atomic clocks, for example, International Atomic Time (TAI) and Terrestrial Time, TT(BIPM)[2] (Petit, 2003).

We make use of two types of statistics for estimating the timekeeping stability: first, the square root Allan variance (SRAV), represented by $\sigma_y(\tau)$ (also called the Allan frequency deviation in more modern literature), which is frequently used in the time and frequency community. The second is the $\sigma_z(\tau)$ statistic (Matsakis *et al.*, 1997), which finds popularity in the pulsar timing community.

The σ_z statistic is used as the quantifier of pulsar time of arrival (TOA) signals that, among other things, are insensitive to the deterministic but *a priori* unknown average period drift $\dot{P}$ (or first derivative of pulse frequency) while allowing one to study the long-term random noise characteristics of the data. This has been found to be suitable when studying millisecond pulsars over the longest time periods—a "pulsar

---

[*]john@physics.uwa.edu.au

[1]University of Victoria, Canada, "Practical Astronomy I: Timekeeping," http://visav.phys.uvic.ca/~babul/AstroCourses/P303/Module1_p3a.htm.

[2]TT(BIPM) is expected to be the most accurate and the most stable time scale over the longest times.

variance.'' The outcome of that analysis has resulted in some pulsars being referred to as ''nature's most stable clocks'' (Taylor, 1991).

We also show that $\sigma_y(\tau)$ with period drift removed and $\sigma_z(\tau)$ yield comparable values over $\tau \approx 10$ year time periods for the best pulsars (measured against atomic clocks). For all periods $\tau < 3$ years, $\sigma_y(\tau)$, with period drift removed, gives a lower noise statistic than does $\sigma_z(\tau)$. Except over the longest measured time periods the pulsar data, shown in this paper, are thus more favorably represented by $\sigma_y(\tau)$.

For the pulsars we have thus removed a measured yet *a priori* unpredictable average period drift $\dot{P}$ from the TOA data, which results in the most optimistic statistic for these pulsars (Verbiest *et al.*, 2009; Verbiest *et al.*, 2008). If a significant level of period drift $\dot{P}$ is present, and not removed before analyzing the random contributions, then it will contribute to the total instability with a $(|\dot{P}|/P)(\tau/\sqrt{2})$ dependence. In the case of modern terrestrial clock-to-clock comparison data it is unnecessary to remove any such effect. And for the very longest averaging times we compare $\sigma_y(\tau)$ and $\sigma_z(\tau)$ for the millisecond pulsar PSR J0437-4715.

The square root Allan variance can be calculated either from the fractional frequency fluctuations of the oscillator or from the period fluctuations in the form of timing residuals, both measured with respect to a lower noise reference. In the 1980s a comparison was made along similar lines (Cordes and Downs, 1985; Il'in *et al.*, 1989) in which it was found that one pulsar, a millisecond pulsar (PSR B1937+21), rivaled the frequency stability of the hydrogen maser (H maser). Those workers concluded, however, that most pulsars were much noisier than a H maser.

There has also been an examination of the potential use of ensemble pulsar timing data to correct the very long-term fluctuations of terrestrial time scales. Based on the quality of the time scales of that era, it was suggested that this could work with the discovery of new high-stability pulsars and longer-term observations (Matsakis and Foster, 1996; Petit and Tavella, 1996). However, since that time there has been a substantial improvement (by up to 2 orders of magnitude) in the accuracy and stability of artificial terrestrial clocks. There have also been new claims that pulsating white dwarfs have a superior stability to even the best millisecond pulsars (Kepler *et al.*, 2005; Mukadam *et al.*, 2003).

Thus, in this paper we hope to present two important additions to the literature: that the improvement in modern terrestrial clock performance means that it is doubtful that astrophysical sources would be able to correct terrestrial time scales on periods less than 10 years, and, in addition, that a reanalysis of the stability of these pulsating white dwarfs actually shows that they are substantially inferior to the millisecond pulsars. We believe it is thus a sensible and opportune time to revisit the claims being made for high quality natural clocks together with a comparison to modern terrestrial clocks.

A. Definition of the SI second

To commence, one should clearly differentiate between the concept of accuracy and stability. On occasion, the word *accuracy* has been loosely applied to the stability of a pulse train delivered by natural clocks such as pulsars (Bronnikov and Mel'nikov, 1989; Karastergiou and Johnston, 2007; Lyne and Graham-Smith, 2006; Thomsen, 1985; Time magazine, 1968; Zhong and Yang, 2007). Any attempt to produce an accurate time standard requires that the frequency of the oscillator can be related to the definition of the SI second, which at the present time is defined in terms of a hyperfine splitting in the ground state of the cesium atom.[3] It is possible that in the future this SI definition may be replaced with another, although any new definition is sure to be closely connected to the values of the universal and presumed unchanging fundamental constants (Taylor and Mohr, 2001).

Contrasting with this universal and unchanging definition, the frequency of a natural astrophysical source is determined by some stochastically distributed initial period together with some additional processes that have occurred in the evolution of the star to its current observed rotational status. It is thus extremely unlikely that these natural pulsating sources could ever be the basis of an accurate time system (Matsakis and Foster, 1996; Matsakis and Josties, 1996; Petit and Tavella, 1996; Riehle, 2004). Furthermore, as we will see below, one can perform a careful accuracy evaluation on terrestrial clocks in order to ascertain all potential sources of systematic uncertainty. This is something that is not possible with the astronomical sources and, as we will show, allows us to extrapolate the frequency stability of the clocks to time periods over which it has not yet been possible to measure. This allows us to place a limit on the worst possible performance in that integration time regime.

B. Pulsar time scale

In spite of this intrinsic inaccuracy, one can still conceive of a pulsar time scale, derived from an agreed-upon ensemble of pulsars, which might prove to be useful if it were to provide a universally accessible and highly *stable* time scale over the longest times (> 10 years) (Matsakis and Foster, 1996; Petit and Tavella, 1996). In light of this possibility it is worthwhile to examine the potential stability of these astrophysical sources. These pulsar timing arrays could potentially verify the extreme long-term performance of terrestrial time scales (Matsakis and Josties, 1996).

Modern artificial clocks have undergone detailed frequency accuracy evaluations[4] at the level of a few parts in

---

[3]SI second: The thirteenth General Conference on Weights and Measures/Conférence Générale des Poids et Mesures (CGPM) (1967/1968, Resolution 1) defined the second to be the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium 133 atom. It follows that the hyperfine splitting in the ground state of the cesium 133 atom is exactly 9 192 631 770 Hz. Then at its 1997 meeting the International Committee for Weights and Measures/Comité International des Poids et Mesures (CIPM) affirmed that this definition refers to a cesium atom at rest at a temperature of 0 K. See www.bipm.org/en/si/si_brochure/chapter2/2-1/second.html.

[4]Technically these clocks cannot be more accurate than cesium primary frequency standards that currently define the second. In order to realize this accuracy it would require a redefinition of the second.

10^{17}. The implication of these evaluations is that the frequency stability will, over some sufficiently long measurement time, exhibit a stability that is at or below this same value. Once the frequency stability has passed under this accuracy evaluation limit, it will not exceed this value for arbitrarily long measurement times. On the other hand, the best pulsars, measured by comparison with local reference clocks, have relatively poor signal to noise leading to a stability of $\sigma_z(\tau) \approx 10^{-15}$ for $\tau \approx 10$ years (Verbiest *et al.*, 2009 and Verbiest *et al.*, 2008). Therefore only by averaging the pulsar pulse timing sequence for times possibly greater than hundreds of years is it possible for a pulsar pulse sequence to present a superior stability to that of the best Earth-based clock. We reiterate that what is currently measured is actually the stability of the pulsar compared to some terrestrial local atomic reference time scale rather than direct pulsar-pulsar comparisons.

This article and analysis is not intended to undermine the potential for pulsars to be excellent tests of cosmology, general relativity, and astrophysics in extreme conditions where they are the most sensitive tests of these type of physics yet performed (Breton *et al.*, 2008; Kramer *et al.*, 2006; Lattimer and Prakash, 2004; Lyne *et al.*, 2004; Manchester, 2004; Stairs, 2004).

II. FREQUENCY STABILITY AND ACCURACY OF CLOCKS

The conventional measure that metrologists use to characterize the stability of any oscillator is the SRAV (Allan, 1966; Barnes *et al.*, 1971). The measure operates on a time series of frequency measurements taken over some integration time τ and is usually quoted over a range of different integration times to capture the behavior of the oscillator over these various time periods. The SRAV is a close relative of the more conventional standard deviation but is stable in the presence of various divergent noise types that afflict real oscillators (C.C.I.R., 1986; Sullivan *et al.*, 1990). In practice the SRAV $\sigma_y(\tau)$ is calculated from a time series of N frequency measurements, $f_n, f_{n+1}, f_{n+2}, \ldots$, where

$$\sigma_y^2(\tau) = \frac{1}{2(N-1)\nu_0^2} \sum_{n=1}^{N-1} (f_{n+1} - f_n)^2, \qquad (1)$$

and ν_0 is the nominal average frequency of the oscillator. The SRAV is defined assuming that the integration time equals the time between samples; for other circumstances there are published corrections (Dawkins *et al.*, 2007). Following Barnes *et al.* (1971) one can also determine the SRAV from

$$\sigma_y^2(\tau) = \frac{1}{8\pi^2\tau^2\nu_0^2(N-2)} \sum_{n=1}^{N-2} [\phi(t_{n+2}) - 2\phi(t_{n+1})$$
$$+ \phi(t_n)]^2, \qquad (2)$$

where $\phi(t)$ is the phase of the oscillator at some epoch t and in this case τ is defined as the time between N phase samples.

In the case of a pulsed source (such as a pulsar) we can rewrite Eq. (2) in terms of timing errors as

$$\sigma_y^2(\tau) = \frac{1}{2\tau^2(N-2)} \sum_{n=1}^{N-2} [X(t_{n+2}) - 2X(t_{n+1}) + X(t_n)]^2, \qquad (3)$$

where $X(t)$ is defined as the time difference (measured in seconds) between the actual epoch of some well-defined event and the predicted epoch of that same event according to some ideal clock.

The subtlety inherent in all the measurement procedures described above is that one must always compare the frequency of some clock under test against another reference clock of equal or better performance if one wishes to be confident about the results of the measurement: This is just as true for natural as it is for artificial clocks.

A. Astrophysical and terrestrial frequency standards compared

The definition for the Allan variance in Eq. (3) is very close to the usual method for measuring the temporal stability of the pulse train emitted by a white dwarf or a pulsar detected on Earth. For example, Kepler *et al.* (2005) measured the instant of the brightest light emission of a white dwarf (G117-B15A) with a 215.197 s period irregularly over ~30 years. We have reanalyzed their data and presented the results in terms of the SRAV in Fig. 1. In addition, on this plot we show
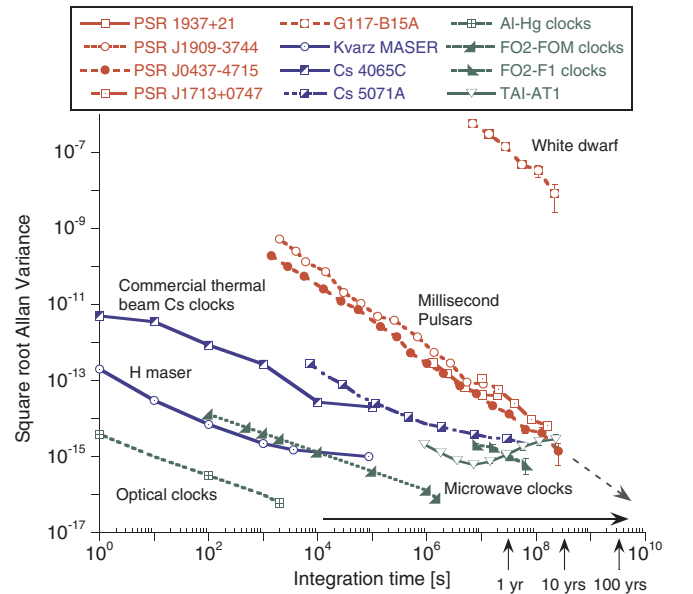

FIG. 1. Derived frequency stability for a range of different frequency sources including astrophysical sources, commercial oscillators, and the best laboratory clocks. See text for a complete description. For clarity, errors have been included only on the most relevant data to the discussion. The solid line arrow indicates the upper limit of the long-term atomic clock stability due to the known accuracy of one of the optical ion clocks at 2.3×10^{-17}. The dashed line arrow shows the long-term trend for the best millisecond pulsar.

σ_y calculated from TOA data for 3 ms pulsars (PSR J1909-3744, PSR J0437-4715, and PSR J1713+0747) (Hotan *et al.*, 2006; Verbiest *et al.*, 2009; Verbiest *et al.*, 2008), as well as one (PSR B1937+21) that was first proposed many years ago as a potential timekeeper (Allan, 1987). The timing residuals used here for PSR J0437-4715 are shown in Verbiest *et al.* (2008) and those for PSR J1909-3744 and PSR J1713+0747 are shown in Verbiest *et al.* (2009). Both of these papers describe in some detail the analysis performed to obtain them, and we have followed exactly the same procedure. The SRAV shown here has been calculated from those timing residuals after linear period drift was removed.

Figure 1 presents the SRAV of a selection of the best artificial clocks and frequency standards derived from measurement procedures as outlined above. The line labeled *Al-Hg clocks* represents a comparison of the Al^+ and Hg^+ ion clocks at NIST, USA, which are both based on optical interrogations of laser cooled and trapped ions. The result reported here is effectively a measurement of the combined stability of both clocks as they have similar stability.

The line labeled *FO2-FOM clocks* shows the best performing microwave frequency standards and is a comparison of a rubidium fountain clock and the mobile cesium fountain from the Observatory of Paris (SYRTE) (Guéna *et al.*, 2008). The line labeled *FO2-F1 clocks* is derived from a 6 year comparison between the atomic fountain clocks at NIST (F1) and at SYRTE (FO2)[5, 6] (T. E. Parker, 2010). The data are somewhat sparse so we do not present stability data beyond 2 years integration time. It is important to be aware that NIST F1 and SYRTE FO2 clocks are similar in performance. The larger instability in the FO2-F1 data comes from the comparison process, due to significant frequency transfer uncertainty, imperfect alignment of runs, and a high rate of dead time (Parker, 2009).

We show the specified frequency stability of a commercial H maser (labeled as Kvarz MASER),[7] which we have confirmed in comparison with a University of Western Australia cryogenic sapphire oscillator (Hartnett *et al.*, 2006). The performances of commercial thermal beam cesium clocks are

shown by the curves labeled Cs 4065C[8] and Cs 5071A.[9] In the latter a particular cesium 5071A standard was compared with the NIST clock ensemble over 7 years until the cesium supply was fully depleted. The time residual data are spaced every 2 hours (with very few missing) from which the SRAV was calculated after linear drift was removed. This represents the entire life of a high-performance Cs tube.

Finally, in the line labeled *TAI-AT1* we show σ_y calculated for the intercomparison of two terrestrial time scales, from the time difference between International Atomic Time (TAI) and a free running NIST clock ensemble (called AT1), after linear drift was removed. We note that these time scales are based on hundreds of Cs atomic clocks and hydrogen masers and that the longevity of a time scale is not dependent upon the type, model, or longevity of any individual device. For example, the gradual introduction of high-performance hydrogen masers into AT1 over the past decade has resulted in an improved stability for that time scale.

The principal feature to note from this plot is the much higher performance of artificial oscillators to clocks in comparison with the astrophysical sources over short measurement times. In addition, in strong contrast with the claims by Kepler *et al.* (2005) the white dwarf stability is substantially inferior when compared with both the millisecond pulsars and the artificial clocks.

It can also be noted that the slopes of the various curves on the plot are not all the same. In the case of the astrophysical sources one sees mostly a slope of $1/\tau$, which is characteristic of a source that is dominated by white phase noise (Rutman, 1978). This is what would be expected in the case of a measurement based on measuring the epoch of a particular event in the presence of white noise (principally receiver noise because of the limited signal to noise ratio).

Most of the atomic clocks show a downward slope of $1/\sqrt{\tau}$, which is characteristic of the presence of white frequency noise (Rutman, 1978) that arises when locking an oscillator to a frequency dispersive feature in the presence of white detection noise. The H-maser and Cs beam clocks do show some evidence of different slopes at the longest times associated with quadratic frequency drift and random walk of frequency.

Although the longest terrestrial measurements trend upwards for the longest integration times to nearly meet the best of the pulsars, one should emphasize that these are known limitations of these older standards. There is no evidence (over their necessarily shorter measurement times) of frequency drift from the SRAV calculated for any of the other modern atomic clocks.

One notes that the FO2-F1 clock comparison has an instability characterized by $\sigma_y \approx 6 \times 10^{-12} \tau^{-1/2}$ over the integration times shown, although, as mentioned, this is not indicative of the clocks themselves but instead due to tech-

---

[5]These data were received 13 May 2009 from Tom Parker at NIST, who did the comparison from data available in Circular T www.bipm.org/jsp/en/TimeFtp.jsp?TypePub=publication#TAIPub FormPubli2 and on the BIPM Web site www.bipm.org/en/scientific/tai/.

[6]Here there were large gaps when both fountains were not running and the fountain runs were also not perfectly aligned. The uncertainty for an individual fountain comparison is made up of the uncertainties of the fountains, the frequency transfer uncertainty for that run, and a dead time uncertainty due to the misalignment of the runs. If the runs are perfectly aligned in time, the dead time uncertainty will be zero, but in many cases this was, in fact, the dominant uncertainty. Thus an uncertainty for an individual comparison is made up of several components. The total comparison uncertainty was typically about 1 or $(1-2) \times 10^{-15}$, and the scatter among the points was consistent with the uncertainties. By calculating the Allan deviation on a series of these points, the noise (no matter how it got into the uncertainties) averages down as $\tau^{-1/2}$ (white frequency noise).

[7]Kvarz SRAV data from Leapsecond.com Web site. See www.leapsecond.com/museum/ch176/.

[8]Symmetricom thermal beam cesium frequency and time standard. See www.gigatest.net/symmetricom TT&M/Precise Frequency References - Atomic Clocks/ds_4065c.pdf. Product now discontinued.

[9]Symmetricom thermal beam cesium frequency and time standard. See http://www.symmetricom.com/products/frequency-references/cesium-frequency-standard/5071A/.

nical issues associated with comparing widely separated clocks on different continents. Similarly pulsars seen from different hemispheres need local reference clocks, and to make pulsar-to-pulsar comparisons of such, the local clock instabilities also affect those measurements.[10] Nevertheless, at its closest approach this clock comparison is still an order of magnitude below the best pulsars, and 3 orders of magnitude better, if we extrapolate the SRAV of the underlying Cs fountain stability (as exemplified by the FO2-FOM clock comparison) so as to eliminate the added uncertainties that resulted from the comparison as explained.

B. Accuracy evaluation

In order to obtain a worst-case estimate of the stability of modern atomic clocks for long-term integration times, which have not been directly measured, we examine recently performed accuracy evaluations on these devices. The accuracy evaluation encapsulates all possible systematic uncertainties of the device and could be interpreted as the potential performance of a time scale if it were to be based on such a device. The accuracy evaluation can also be regarded as the maximum fractional deviation of the clock output frequency from the "true" underlying and unperturbed atomic transition frequency. Thus one sees that, given sufficient integration to average away white noise associated with the measurement process, it is clear that the frequency instability, measured between two clocks with a certain accuracy evaluation, cannot exceed the bound set by the accuracy evaluation itself.

The most recent accuracy evaluations have been performed for the Hg and Al optical ion clocks (Rosenband *et al.*, 2008) at the level of 1.9×10^{-17} and 2.3×10^{-17}, respectively, the Sr optical clock (Akatsuka *et al.*, 2008; Campbell *et al.*, 2008; Ludlow *et al.*, 2008) at $(1–1.4) \times 10^{-16}$, the Yb optical clock (Lemke *et al.*, 2009) at 3.4×10^{-16}, and the Cs and Rb fountain clocks at SYRTE (Chapelet *et al.*, 2007; Guéna *et al.*, 2008) at the level of about 4.5×10^{-16}. It should be noted that the Hg and Al optical ion clock accuracy evaluations are below that of the current realization of the SI second; hence, they have actually reported the ratio of their optical frequencies, so as to avoid the uncertainty (3.3×10^{-16}) of the currently realized SI second (Jefferts *et al.*, 2007). The same is also true for the Sr lattice clock in which a remote optical comparison was made with an atomic Ca clock (Ludlow *et al.*, 2008).

As an example of the power of the systematic uncertainty evaluation, the particular Cs 5071A beam clock used to make the comparison, shown as Cs 5071A in Fig. 1, has an accuracy[11] (T. Parker, 2010) of 2×10^{-15}, and one sees that its

stability approaches that value at the longest times. One can also see that the stability of the Al-Hg ion clocks comparison also approaches its measured accuracy evaluation at the longest integration times, indicated by a solid arrow in Fig. 1. We note that commercial Cs beam atomic clocks may be more stable over 1 year of averaging than over 5 years, but nonetheless their long-term stability always remains below their specified systematic uncertainty (which sets the upper bound for the long-term stability).

The implication of this analysis is that the extrapolated worst-case Hg or Al ion clock stability (essentially its systematic uncertainty) will equal the stability of the extrapolated best millisecond pulsar stability at an integration time of the order of hundreds of years (10^{10} s): a measurement that would take many thousands of years of integration time to properly characterize.

The dimensionless ratio of two ideal unperturbed transition frequencies of different ions or atoms should be constant in the same reference frame (whether inertial or not) using the current standard model of particle physics (Uzan, 2003). The ideal atomic clock attempts as closely as possible to yield an output frequency equal to that of an unperturbed transition, and the accuracy evaluation is a combined theoretical and experimental test to quantify the difference between the actual output and the unperturbed value. It is, of course, always possible that there is some currently unknown physics beyond the standard model that will further perturb the transition, and it is indeed this possibility that motivates current searches for failures in the standard model of physics using clock comparisons and atom interferometers. It is only meaningful to look for new physics in the variation of dimensionless ratios (Flambaum, 2007; Flambaum and Kozlov, 2007; Uzan, 2003). To date, however, all tests aimed at searching for such new physics have yielded results consistent with the existing accuracy limits. Nonetheless, one can accept that extreme long-term comparisons of time scales provided by ensembles of pulsars against that provided by atomic clocks might yield hints of some particular flaw in the current model of physics, which is not evident in comparisons of various atomic clocks, but which is seen because of the different physics that dominates the astrophysical sources.

III. PERIOD DRIFT

It appears that Kepler *et al.* (2005) have confused the period drift rate $\dot{P}/P$ with the fractional variance of the period [which is numerically equal to the fractional variance of the frequency as defined in Eq. (3), i.e., $\dot{P}(t)/P(t) = -\dot{f}(t)/f(t)$]. When using the corrected analysis presented here, it is interesting to revisit the comparison of the drift rates of the pulse-to-pulse period of astrophysical sources against the drift rates of the best artificial clocks.

In the circumstances where random walk of frequency is absent it is possible to obtain a stable estimate of the drift by fitting a line to a time sequence of frequency measurements (Barnes *et al.*, 1971). We have obtained these data for the best modern atomic clocks from measurements made in connection with tests of the temporal stability of the universal constants (Fortier *et al.*, 2007; Guéna *et al.*, 2008; Rosenband *et al.*, 2008) and summarized these in Table I.

---

[10]We note proposals to monitor many pulsars simultaneously, which can eliminate this dependence on terrestrial time scales, or at least provide a flywheel to correct atomic time over very long time periods.

[11]At NIST they are able to calibrate the Cs 5071A clocks against the NIST-F1 standard to an accuracy better than 1×10^{-15}; hence the accuracy evaluation for the particular cesium standard used here was estimated to be 2×10^{-15}. This is much better than the accuracy quoted by the manufacturer, without such calibration, usually about 5×10^{-13}.

TABLE I.   Clocks and astrophysical sources and their period drift rates

Source	$\dot{P}/P$ (s^{-1})	Reference
H^{199}g$^+$ ion vs Cs clock	1.3×10^{-23}	Fortier *et al.* (2007)
Y^{171}b$^+$ vs Cs clock	1.4×10^{-22}	Peik *et al.* (2004)
H $1S - 2S$ vs Cs clock	2.0×10^{-22}	Fischer *et al.* (2004)
PSR B1937+21	6.74×10^{-17}	Rawley *et al.* (1987)
PSR J0437-4715	9.95×10^{-18}	Hotan *et al.* (2006)
PSR J1909-3744	4.76×10^{-18}	Hotan *et al.* (2006)
PSR J1713+0747	1.86×10^{-18}	Hotan *et al.* (2006)
G117-B15A	1.66×10^{-17}	Kepler *et al.* (2005)
ZZ Ceti, R548	2.58×10^{-17}	Mukadam *et al.* (2003)

It should be noted that these drift rates are obtained from the 1σ uncertainty of drift rates that are consistent with zero. Hence it can be seen that the upper limit of these drift rates for the best artificial clocks are 5 orders of magnitude below the lowest values for the pulsars.

The experimental measurement of the stability of white dwarf G117-B15A (Kepler *et al.*, 2005) should have suggested that the performance of the star as a clock was far from the state of the art. Kepler *et al.* indicated that they measured the epoch of the maximum pulse brightness to a resolution of the order of a second and then compared with the time delivered by a local time dissemination system (atomic time scale). As can be seen from Eq. (3), no normalization to period should be made in calculating the SRAV from timing errors, and keeping time to an accuracy of a second (even over 30 years) is far from the state of the art in modern time dissemination systems. Since the performance of the white dwarf pulses is inferior to that of the time dissemination system, it also implies that the stability and drift are also worse than the local time reference.

IV. σ_y AND σ_z OVER THE LONGEST MEASUREMENT TIMES

Power law fits to the calculated SRAV for the three pulsars in Fig. 1 have slopes very close to $1/\tau$, indicating white phase noise over these time scales. The noise is dominated by signal to noise considerations in the receiver and time delays in the electronics, instrumental polarization, propagation delays in the interstellar plasma, and pulse phase jitter intrinsic to the pulsar. It is possible that this can be improved somewhat by averaging longer (Hotan *et al.*, 2006) and significant improvements in TOA precision can be made by paying close attention to the above mentioned effects, including improvements to receiver systems and construction of telescopes with larger collecting areas.

It has recently been reported (Akatsuka *et al.*, 2008; Oates, 2008), however, that in the relatively near future optical lattice neutral atom clocks and trapped ion clocks could potentially demonstrate a frequency stability of the order of $10^{-17}/\sqrt{\tau}$ were they to be able to overcome a challenging technical issue associated with their interrogating oscillators. This would make them more than 2 orders of magnitude better than the best terrestrial clocks shown in Fig. 1.

Given that the best clocks and best pulsar observatories are unlikely to be co-located, it is perhaps better to consider that the uncertainties exemplified by the comparison of fountain clocks located on different continents (FO2-F1 in Figs. 1 and 2) are the fairest representation of the level of performance to be expected from a state-of-the-art clock with a high quality time transfer system. The long-term performance of this complete system is limited by the systematic uncertainty of the Cs fountains, which is at such a level that the stability of the astrophysical sources could only exceed that of the joint clock-time transfer system for periods beyond a few hundred years of integration time.

These clocks (FO2 and F1) were operated neither continuously nor simultaneously. Their comparison is based on individual comparisons with TAI, each lasting 15 to 30 days, and scattered during the nearly 11 year existence of the NIST F1 fountain. With the known properties of the underlying reference TAI one can indeed build a long-term comparison. And, since time transfer noise averages much faster than white frequency noise, the time transfer noise is well below the limit of current state-of-the-art terrestrial clocks. However, the uncertainties associated with intercontinental comparisons still present a challenge, and most re-
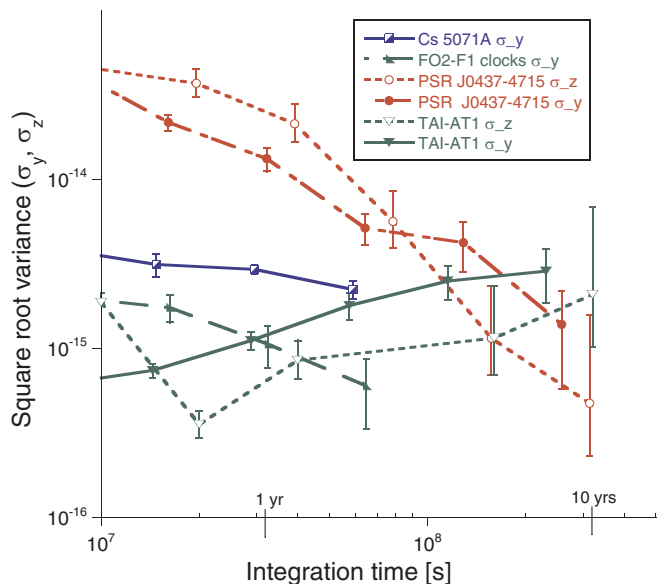

FIG. 2.   Derived frequency stability for some of the frequency sources of Fig. 1 for comparison over the longest time scales. Included are the pulsar PSR J0437-4715, a commercial cesium oscillator Cs 5071, and some of the best laboratory clocks. The TAI-AT1 time scale comparison data and the pulsar-to-clock comparison data are represented with both the σ_y and σ_z statistics.

searchers believe that recent work on time and frequency dissemination using optical fiber networks (Daussy *et al.*, 2005; Kéfélian *et al.*, 2009) and comparisons to space-borne clocks (Uhrich *et al.*, 2000) will result in orders of magnitude improvement to long-baseline clock comparisons in the very near future.

Finally a closer look is needed over the very longest times. For this reason we have expanded the scale of Fig. 1 around integration times of 1 to 10 years and compare only the best pulsar (which was measured against a terrestrial time scale) with terrestrial clock-to-clock and time scale-to-time scale comparisons that have data over that time period. This is shown in Fig. 2.

Data from the TAI-AT1 time scale comparison and the millisecond pulsar PSR J0437-4715 are presented both as σ_z and σ_y, the latter with period drift removed. The last σ_y point for the PSR J0437-4715 data is determined from one-half of the data set averaged with the other half: it is for this reason that the error bars are so large. The final σ_z point for the PSR J0437-4715 pulsar and the TAI-AT1 time scale data comes from the length of the whole times series data set. In the case of the pulsar that is 9.9 years (Verbiest *et al.*, 2009) and in the case of the terrestrial time scales comparison, the last 10 years of available data were used in the analysis. That time period corresponds to the same period of the pulsars data. The computed 1σ error bars come from the limits placed on σ_z^2 assuming it has a χ^2 distribution with n degrees of freedom, where n is the number of squared values of c_3 (the coefficient of the best fit cubic term in the time series data) appearing in the average. However, for the final few values the estimates of the σ_z statistic will be biased low (Matsakis *et al.*, 1997).

It should be noted that for the longest measured averaging times, σ_y and σ_z are essentially equivalent when the error bars are considered. It is clear that the quality of the data for this pulsar (which was measured against some terrestrial time scale) is insufficient to really separate the two measures. One sees that for $\tau > 6$ years σ_y of the pulsar and the TAI-AT1 comparison are also equivalent. This has to be the case as the pulsar measurements are compared to a disseminated time scale as well, and so the results must be similar. The uncertainty in the time scale TT(BIPM2003), used by pulsar timing researchers, is 1×10^{-15} (Petit, 2003). The apparent slight differential between the final σ_z point for the pulsar and the final σ_y values can be explained because of the increased rejection of the σ_z measure of long-term noise and the non-stationarity of the noise over these long-term comparisons [see Fig. 5 in Matsakis and Foster (1996)]. This is emphasized by the final two σ_z values for both the pulsar and the TAI-AT1 time scale comparison; within the error bars they are equivalent.

V. CONCLUSION

In conclusion, previous publications have indicated that some pulsars, or even white dwarfs, could, or do, provide the best clocks in the Universe. Although measurements on time periods beyond 7 years have yet to be performed on the most advanced terrestrial clocks, and there are insufficient data to properly characterize them beyond 2 years, the above analysis

strongly implies that artificial clocks constructed within the time and frequency community are very likely to be better in terms of accuracy or stability over any accessible time period.

Accuracy evaluations (Chapelet *et al.*, 2007; Guéna *et al.*, 2008; Lemke *et al.*, 2009; Rosenband *et al.*, 2008) on modern clocks imply that with longer measurements their stability will get to at least the parts-in-10^{17} regime. Natural clocks, on the other hand, would require measurement times over many hundreds of years before they could confidently claim a frequency stability of this order. The trend line for σ_y in Figs. 1 and 2 or even for σ_z in Fig. 2 indicates that it would require an ensemble (to eliminate short-term nondeterministic noise) of pulsars averaged over many hundreds of years to reach an rms stability level of parts in 10^{17}, which is an upper limit on the most accurate terrestrial clocks.

The proviso on this conclusion is that advanced terrestrial clocks have not yet been measured over periods greater than about 6 years, which means statistics for periods greater than 2 years are not yet available, nor have time dissemination systems operated that will allow this stability to be broadly available globally (something that is equally crucial for comparison of astrophysical sources using these advanced terrestrial clocks and time scales including them). On the latter point we note active development in this field has already successfully demonstrated sub-10^{-15} accuracy transfer over hundreds of kilometers (Kéfélian *et al.*, 2009). On the former point we note that the accuracy and stability of terrestrial clocks have improved more than an order of magnitude, on average, in each decade over the past 60 years, since the development of the first atomic Cs clock, while the timing stability of the best millisecond pulsar detected by comparison with local atomic clocks has improved by less than an order of magnitude in the past 20 years[12] (Petit, 2010). Recently identified paths (Akatsuka *et al.*, 2008; Oates, 2008) to further improved terrestrial clocks suggests that this trend will continue into the future.

REFERENCES

Akatsuka, T., M. Takamoto, and H. Katori, 2008, Nature Phys. **4**, 954.

Allan, D. W., 1966, Proc. IEEE **54**, 221.

Allan, D. W., 1987, in *Proceedings of the 41st Annual Symposium on Frequency Control* (IEEE Publishing Services, New York), p. 2, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1537934.

Allan, D. W., M. A. Weiss, and T. K. Peppler, 1989, IEEE Trans. Instrum. Meas. **IM-38**, 624.

Barnes, J., A. Chi, L. Cutler, D. Healey, D. Leeson, T. McGunigal, J. Mullen, W. Smith, R. Sydnor, R. Vessot, and G. Winkler, 1971, IEEE Trans. Instrum. Meas. **IM-20**, 105.

Breton, R. P., V. M. Kaspi, M. Kramer, M. A. McLaughlin, M. Lyutikov, S. M. Ransom, I. H. Stairs, R. D. Ferdman, F. Camilo, and A. Possenti, 2008, Science **321**, 104.

Bronnikov, K. A., and V. N. Mel'nikov, 1989, Izmer. Tekh. **1**, 29 [Meas. Tech. **32**, 46 (1989)].

Campbell, G. K., A. D. Ludlow, S. Blatt, J. W. Thomsen, M. J. Martin, M. H. G. de Miranda, T. Zelevinsky, M. M. Boyd, Jun Ye, S. A. Diddams, T. P. Heavner, T. E. Parker, and S. R. Jefferts, 2008, Metrologia **45**, 539.

C.C.I.R., 1986, International Radio Consultative Committee Report 580, pp. 142–150.

Chapelet, F., J. Guéna, G. D. Rovera, P. Laurent, P. Rosenbush, G. Santarelli, S. Bize, and A. Clairon, 2007, in *Proceedings of the Joint 21st EFTF and IEEE Frequency Control Symposium*, Geneva, Switzerland, p. 467.

Cordes, J. M., and G. S. Downs, 1985, Astrophys. J. Suppl. Ser. **59**, 343.

Daussy, C., O. Lopez, A. Amy-Klein, A. Goncharov, M. Guinet, C. Chardonnet, F. Narbonneau, M. Lours, D. Chambon, S. Bize, A. Clairon, G. Santarelli, M. E. Tobar, and A. N. Luiten, 2005, Phys. Rev. Lett. **94**, 203904.

Dawkins, S. T., J. J. McFerran, and A. N. Luiten, 2007, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 918.

Fischer, M., N. Kolachevsky, M. Zimmermann, R. Holzwarth, T. Udem, T. W. Hänsch, M. Abgrall, J. Grünert, I. Maksimovic, S. Bize, H. Marion, F. P. Dos Santos, P. Lemonde, G. Santarelli, P. Laurent, A. Clairon, C. Salomon, M. Haas, U. D. Jentschura, and C. H. Keitel, 2004, Phys. Rev. Lett. **92**, 230802.

Flambaum, V. V., 2007, Int. J. Mod. Phys. A **22**, 4937.

Flambaum, V. V., and M. G. Kozlov, 2007, Phys. Rev. Lett. **99**, 150801.

Fortier, T. M., N. Ashby, J. C. Bergquist, M. J. Delaney, S. A. Diddams, T. P. Heavner, L. Hollberg, W. M. Itano, S. R. Jefferts, K. Kim, F. Levi, L. Lorini, W. H. Oskay, T. E. Parker, J. Shirley, and J. E. Stalnaker, 2007, Phys. Rev. Lett. **98**, 070801.

Guéna, J., F. Chapelet, P. Rosenbusch, P. Laurent, M. Abgrall, G. D. Rovera, G. Santarelli, S. Bize, A. Clairon, and M. E. Tobar, 2008, in *Proceedings of the 2008 IEEE International Frequency Control Symposium* (IEEE, New York), p. 366.

Hama, S., M. Imae, H. Kiuchi, H. Takaba, and F. Takahashi, 1989, in *Proceedings of the 21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, Vol. 21, p. 289, http://www.pttimeeting.org/archivemeetings/1989papers/Vol 21_27.pdf.

Hartnett, J. G., C. R. Locke, E. N. Ivanov, M. E. Tobar, and P. L. Stanwix, 2006, Appl. Phys. Lett. **89**, 203513.

Hotan, A. W., M. Bailes, and S. M. Ord, 2006, Mon. Not. R. Astron. Soc. **369**, 1502.

Il'in, V. G., Y. P. Ilyasov, A. D. Kuz'min, S. B. Pushkin, Y. P. Shitov, and T. V. Shabanova, 1989, Izmer. Tekh. **4**, 25 [Meas. Tech. **32**, 323 (1989)].

Jacoby, B. A., 2005, Ph.D. thesis (CalTech, Pasadena, CA).

Jefferts, S. R., T. P. Heavner, T. E. Parker, and J. H. Shirley, 2007, Proc. SPIE Int. Soc. Opt. Eng. **6673**, 667309, available at http://tf.nist.gov/timefreq/general/pdf/2254.pdf.

Karastergiou, A., and S. Johnston, 2007, Mon. Not. R. Astron. Soc. **380**, 1678.

Kéfélian, F., O. Lopez, H. Jiang, C. Chardonnet, A. Amy-Klein, and G. Santarelli, 2009, Opt. Lett. **34**, 1573.

Kepler, S. O., J. E. S. Costa, B. G. Castanheira, D. E. Winget, F. Mullally, R. E. Nather, M. Kilic, T. von Hippel, A. S. Mukadam, and D. J. Sullivan, 2005, Astrophys. J. **634**, 1311.

Kramer, M., I. H. Stairs, R. N. Manchester, M. A. McLaughlin, A. G. Lyne, R. D. Ferdman, M. Burgay, D. R. Lorimer, A. Possenti, N. D'Amico, J. M. Sarkissian, G. B. Hobbs, J. E. Reynolds, P. C. C. Freire, and F. Camilo, 2006, Science **314**, 97.

Lattimer, J. M., and M. Prakash, 2004, Science **304**, 536.

Lemke, N. D., A. D. Ludlow, Z. W. Barber, T. M. Fortier, S. A. Diddams, Y. Jiang, S. R. Jefferts, T. P. Heavner, T. E. Parker, and C. W. Oates, 2009, Phys. Rev. Lett. **103**, 063001.

Ludlow, A. D., T. Zelevinsky, G. K. Campbell, S. Blatt, M. M. Boyd, M. H. G. de Miranda, M. J. Martin, J. W. Thomsen, S. M. Foreman, Jun Ye, T. M. Fortier, J. E. Stalnaker, S. A. Diddams, Y. Le Coq, Z. W. Barber, N. Poli, N. D. Lemke, K. M. Beck, and C. W. Oates, 2008, Science **319**, 1805.

Lyne, A. G., M. Burgay, M. Kramer, A. Possenti, R. N. Manchester, F. Camilo, M. A. McLaughlin, D. R. Lorimer, N. D'Amico, B. C. Joshi, J. Reynolds, and P. C. C. Freire, 2004, Science **303**, 1153.

Lyne, A. G., and F. Graham-Smith, 2006, *Pulsar Astronomy*, 3rd ed. (Cambridge University Press, Cambridge, England) pp. 62, 63.

Manchester, R. N., 2004, Science **304**, 542.

Matsakis, D. N., and R. S. Foster, 1996, in *Amazing Light*, edited by R. Y. Chiao (Springer Verlag, Berlin).

Matsakis, D. N., and F. J. Josties, 1996, in *Proceedings of the 28th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, Vol. 28, p. 225, http://www.pttimeeting.org/archivemeetings/1996papers/Vol 28_19.pdf.

Matsakis, D. N., J. H. Taylor, and T. M. Eubanks, 1997, Astron. Astrophys. **326**, 924.

Mukadam, A. S., S. O. Kepler, D. E. Winget, R. E. Nather, M. Kilic, F. Mullally, T. von Hippel, S. J. Kleinman, A. Nitta, J. A. Guzik, P. A. Bradley, J. Matthews, K. Sekiguchi, D. J. Sullivan, T. Sullivan, R. R. Shobbrook, P. Birch, X. J. Jiang, D. W. Xu, S. Joshi, B. N. Ashoka, P. Ibbetson, E. Leibowitz, E. O. Ofek, E. G. Meistas, R. Janulis, D. Alisauskas, R. Kalytis, G. Handler, D. Kilkenny, D. O'Donoghue, D. W. Kurtz, M. Muller, P. Moskalik, W. Ogloza, S. Zola, J. Krzesinski, F. Johannessen, J. M. Gonzalez-Perez, J-E. Solheim, R. Silvotti, S. Bernabei, G. Vauclair, N. Dolez, J. N. Fu, M. Chevreton, M. Manteiga, O. Suarez, A. Ulla, M. S. Cunha, T. S. Metcalfe, A. Kanaan, L. Fraga, A. F. M. Costa, O. Giovannini, G. Fontaine, P. Bergeron, M. S. O. O'Brien, D. Sanwal, M. A. Wood, T. J. Ahrens, N. Silvestri, E. W. Klumpe, S. D. Kawaler, R. Riddle, M. D. Reed, and T. K. Watson, 2003, Astrophys. J. **594**, 961.

Oates, C., 2008, Nature Phys. **4**, 910

Parker, T., personal communication on 15 May 2009

Parker, T., email communication on 9 February 2010.

Parker, T. E., 2010, Metrologia **47**, 1 (Fig. 3 therein for time series data).

Peik, E., B. Lipphardt, H. Schnatz, T. Schneider, C. Tamm, and S. G. Karshenboim, 2004, Phys. Rev. Lett. **93**, 170801.

Petit, G., 2003, in *Proceedings of the 35th Annual Precise Time and Time Interval (PTTI) Meeting*, Vol. 35, p. 307, www.pttimeeting.org/archivemeetings/2003papers/paper29.pdf.

Petit, G., 2010, Talk entitled "Millisecond pulsars to transfer the accuracy of atomic time," European Frequency and Time Forum, Noordwijk, Netherlands, 13–16 April 2010.

Petit, G., and P. Tavella, 1996, Astron. Astrophys. **308**, 290.

Rawley, L. A., J. H. Taylor, M. M. Davis, and D. W. Allan, 1987, Science **238**, 761.

Riehle, F., 2004, *Frequency Standards: Basics and Applications* (Wiley-VCH Verlag GmbH Co., Weinheim), pp. 419–420.

Rosenband, T., D. B. Hume, P. O. Schmidt, C. W. Chou, A. Brusch, L. Lorini, W. H. Oskay, R. E. Drullinger, T. M. Fortier, J. E. Stalnaker, S. A. Diddams, W. C. Swann, N. R. Newbury, W. M.

Itano, D. J. Wineland, and J. C. Bergquist, 2008, Science, **319**, 1808.

Roy, J., T. L. Venkatasubramani, and B. C. Joshi, 2005, in *Proceedings of the XXVIIIth URSI General Assembly, New Delhi, India*, http://www.ursi.org/Proceedings/ProcGA05/pdf/A06.2(01130).pdf.

Rutman, J., 1978, Proc. IEEE **66**, 1048.

Sazhin, M. V., 1989, Izmer. Tekh. **1**, 27 [Meas. Tech. **32**, 42 (1989)].

Stairs, I. H., 2004, Science **304**, 547.

Sullivan, D. B., D. W. Allan, D. A. Howe, and F. L. Walls, Eds., March 1990, U.S. Department of Commerce, National Institute of Standards and Technology, NIST Technical Note 1337.

Taylor, B. N., and P. J. Mohr, 2001, IEEE Trans. Instrum. Meas. **50**, 563.

Taylor, J. H., 1991, Proc. IEEE **79**, 1054.

Thomsen, D. E., 29 June 1985, http://findarticles.com/p/articles/mi_m1200/is_v127/ai_3833111/.

Time magazine, 1968; see www.time.com/time/magazine/article/0,9171,900146,00.html.

Uhrich, P. J.-M., P. Guillemot, P. Aubry, F. Gonzalez, and C. Salomon, 2000, IEEE Trans. Ultrason. Ferroelectr. Freq. Control, **47**, 1134.

Uzan, J-P., 2003, Rev. Mod. Phys. **75**, 403.

Verbiest, J. P. W., M. Bailes, W. A. Coles, G. B. Hobbs, W. van Straten, D. J. Champion, F. A. Jenet, R. N. Manchester, N. D. R. Bhat, J. M. Sarkissian, D. Yardley, S. Burke-Spolaor, A. W. Hotan, and X. P. You, 2009, Mon. Not. R. Astron. Soc. **400**, 951.

Verbiest, J. P. W., M. Bailes, W. van Straten, G. B. Hobbs, R. T. Edwards, R. N. Manchester, N. D. R. Bhat, J. M. Sarkissian, B. A. Jacoby, and S. R. Kulkarni, 2008, Astrophys. J. **679**, 675.

Zhong, C., and T. Yang, 2007, Chin. Astron. Astrophys. **31**, 443.