# Improved Timing for better performance of distributed data acquisition systems

Alexandru Marius PANAIT*

*Corresponding author
INCAS – National Institute for Aerospace Research "Elie Carafoli",
B-dul Iuliu Maniu 220, Bucharest 061126, Romania,
panait.marius@incas.ro

DOI: 10.13111/2066-8201.2021.13.1.12

*Received: 03 December 2020/ Accepted: 26 January 2021/ Published: March 2021*


***Abstract:*** *Distributed data acquisition systems are the norm in the great majority of industry branches where the process to be controlled covers a physically large and often fragmented area. Many local smaller data acquisition modules are interfaced and pass raw or pre-processed data along with the timing signals. These timing signals are a vital part of data acquisition as oftentimes raw data is processed as time series and correlations are made based on timestamps. This paper presents a study for implementation of a cost-effective high precision time base (reference clock) using a GPS receiver as a primary time source and a local high precision real time clock as a secondary oscillator.*

***Key Words****: time synchronization, distributed data acquisition systems, GPS DO, high precision timer*

## 1. TIMING AND DATA AS TIME SERIES

In many industrial applications raw data is processed as a time series. Data contains raw values paired with time stamps, useful for synchronization between data sources and for frequency and correlation analysis. Sometimes timestamps are implicit and other times apparent but the basis of accurate signal representation, processing and analysis is a stable time base. Complex protocols such as EtherCat but also simpler ones such as I2C all define a way to incorporate explicitly or implicitly time in the transported and encoded data.

Timing is therefore an essential part of every data acquisition system. For all their complexity, all the temporal data management implementations fall in two categories: the centrally synchronized master-clock -driven scheme and the distributed "consensus of all subsystems" mean pulse technique.

The first requires a dedicated time signal and an associated protocol while the latter can have time information simply mixed with "normal" data packets traffic. Hybrid schemes do exist and usually consist in a high precision master clock and a network server and NTP or SNTP protocols that distribute the time information to networked data acquisition systems. In any case, the basis of such systems is a reliable timing source, either local or locally managed remote located.

Synchronization is handled differently based on the type of device and whether it is networked or not. For all variants, a basic ultra-stable clock source is useable either directly or by passing it to a timing server on the dedicated network.

Timing errors mostly represented by jitter are responsible for signal distortion and are a major identified source of measurement error [1] so accurately synchronizing multiple data sources and ensuring a stable time base for the data acquisition system is one of the primary concerns in any serious data collection activity.

Jitter is defined as medium and high frequency deviations of the clock signal while wander is used to describe low frequency variations.

Since most data collection is made under the assumption of a stable and known data rate, these distortions induce amplitude and phase distortion in the signals [1] and make accurate signal reconstruction impossible or highly problematic at least and throws correlated signals out of sync.

Timing jitter causes random occurrence of waveform samples and can be either individual, sample jitter or frame jitter in the case of trigger timing errors in triggered sample data acquisition.

These two types of jitter cause inaccurate averaging and errors in signal power measurements [1], [2] by inducing incoherence either between individual samples or between consecutive frames, or between previously synced signals.

Of course, as the time base is distorted, time interval measurement is also affected [1], [2], [3]. A large collection of data sampled with frame jitter and analyzed is presented in [4] where jitter trends are also discussed.

Frame timing errors (frame jitter) can introduce errors as beautifully illustrated by [1], where the sum of two sampled phase-opposed sinusoids is supposed to be zero but if measurements are affected by frame jitter and thus have an induced phase difference then the apparent result is non-zero, and its magnitude and phase is dictated by the noise type and cause. Frame noise induces a continuous difference signal while single sample jitter can induce isolated spikes with noise profiles linked to the time base distortion.

The simplified sampling convention (1) uniquely associates a signal sample $x[i]$ extracted from the continuous time-varying signal $x(t)$ with an ordinal index $i$, forming an ordered string that results as a consequence of multiplication of the initial function $x(t)$ with a uniformly spaced Dirac comb or Shah function:

$$(Ш_T x)(t) = \sum_{k=-\infty}^{\infty} x(t)\delta(t-kT) = \sum_{k=-\infty}^{\infty} x(kT)\delta(t-kT) = \sum_{i=0}^{\infty} x_i \tag{1}$$

*where* Ш*(T) is the periodic Shah function or Dirac comb with period T*

*$x(t)$ is the sampled time-varying signal,*

*$x(i)$ is the ordered sample that defines the sampled signal.*

The Shah function or Dirac comb is defined as shown in (2):

$$Ш_T(t) = \sum_{k=-\infty}^{\infty} \delta(t-kT) = \frac{1}{T}\,Ш\left(\frac{t}{T}\right) \tag{2}$$

where $T$ = Dirac comb function period, that defines the sampling interval.

The comb function in (2) is actually not a proper mathematical function but a generalized function, a periodic tempered distribution built from repeating the Dirac delta function.

The period conventionally marked $T$ defines the non-zero values of the Shah function, and the points where the function has a normalized value of 1 are conventionally named nodes. Multiplication of a function with a Dirac delta periodic distribution results in equidistant samples with the interval $T$.

Their integrals are equal to the value of the sampled function at the nodes of the Dirac comb. Periodization of aperiodic signals can also be modelled by convolution with the Dirac comb [5]. The Dirac comb has the property of Fourier invariance, that is to say that

$$Ш_T\,(t) \overset{\mathcal{F}}{\rightarrow} \frac{1}{T}\,Ш_{\frac{1}{T}}\,(f) \tag{3}$$

or that the Dirac comb Fourier transform remains a Dirac comb.
The Fourier analysis of the above result leads to the well-known Shannon sampling theorem like shown in (4) [6]:

$$Ш_T\,x \overset{\mathcal{F}}{\rightarrow} \frac{1}{T}\,Ш_{\frac{1}{T}} * X \tag{4}$$

But convolution of a function f with a delta function δ (*t*-*kT*) simply means shifting the function by *kT,* so convolution with a Dirac comb is equivalent to a periodic sum such as (5):

$$\left(Ш_{\frac{1}{T}} * X\right)(f) = \sum_{k=-\infty}^{\infty} X(f - \frac{k}{T}) \tag{5}$$

and if we consider a frequency interval (-*a*, *a*) where the function has non-zero spectral components, i.e. a limited frequency domain, this means that only using samples of the original signal at $\frac{1}{2a}$ intervals are sufficient for reconstructing the signal proper as by multiplying the sampled function by a rectangle function Π like in (6) we get the original signal in the time domain [6]:

$$\frac{1}{2a}\prod\left(\frac{t}{2a}\right)(2aШ_{2a} * X) = X \tag{6}$$

where the term $2aШ_{2a} * X$ is the Fourier transform of the convolution of the function x with a Dirac comb of period $\frac{1}{2a}$.

This is the so-called “sampling theorem” and it is obvious from its form that it needs a fixed sampling frequency a to carry out the signal recompositing in (6).

This short formalism helps understand the basis of all fixed data rate sampling commonly used in scientific data collection and explains why a variance in sampling rate is equivalent to spurious data and phase and amplitude distortions.

Single element jitter is equivalent to different sampling times for each affected sample, that is instead of having a single equation (6) we get a sum of several expressions of the same form with different 1/2a = f factor, so that the composed signal is a sum of inverted Fourier transforms (4), leading to the formation of a signal plus noise construct.

When working with raw values though the two are no longer separable as the wrong assumption is made in equation (6) by defining a single window function Π(t/2a) for the intended sampling frequency and the error is carried over at the reconstitution of the signal from its samples.

Aliasing thus ensues with a plethora of other adverse phenomena, such as phase errors. The time series, assumed to be with a fixed data rate of 1/T = f is compromised, as samples from different data rate series are mixed in it.

Distortions ensue when the samples are equally spaced and integrated to get the original function. If a precise oscillator was present in the system, the data could have been recovered using another channel to sample simultaneously this fixed and precise pulse and align the sampled data points to the sampled timer pulses, as in variable bit rate encoding schemes.

The problem of scatter free sampling is reduced to a problem of equal scatter on all inputs of the data acquisition system (easier to achieve) as shown in [1]. In the following chapters all references to time series will describe strings of fixed data rate points sampled from a real-world signal.

Figure 1 shows a properly sampled signal (a simple sine, for convenience) and the effects of scatter and jitter on sampling (or, i.e., desynchronization in distributed data acquisition systems) are shown in figure 2 below.

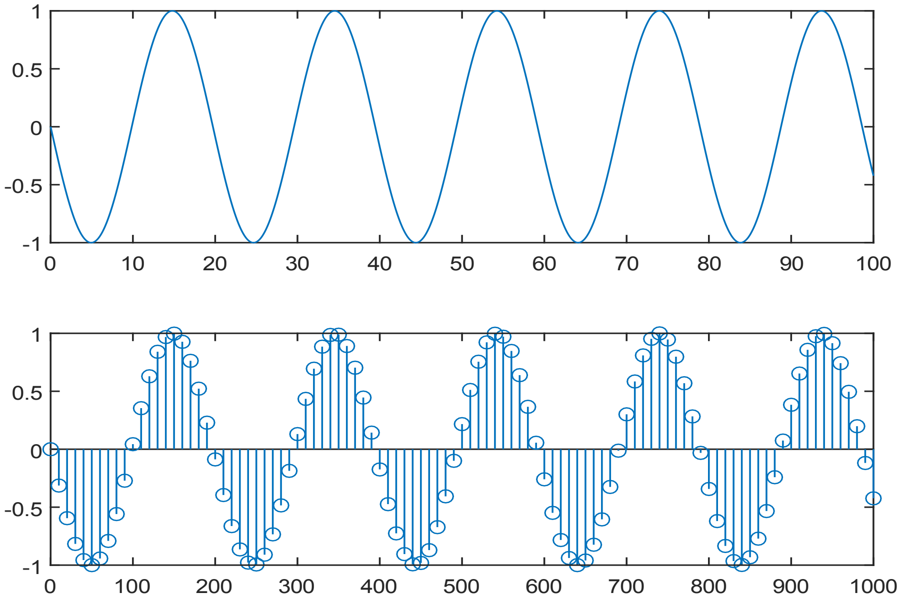


Figure 1. A properly sampled signal is shown here, with the original signal in the subplot above and the stem plots below representing the sampling points. The sampling is correct with regard to the Shannon theorem and shows no jitter or scatter.

Observe the phase difference that appears as a consequence of the jitter in Figure 2. Here small clock errors were simulated by introducing a random timing error generated by using a random number generator and adding as decimals to normally generated sample time vector points.

The errors are small at about 1%, and simulate a naturally occurring time stamp jitter common in most data acquisition systems with free-running local clocks (simple or temperature-compensated crystal oscillators).

But although the waveform is recognizable (the familiar shape of a sine was chosen for clarity), the time distortion induces random phase errors from the actual signal. Phase coherence is lost- both self-coherence (distortion of the signal itself) but also signal-to-signal coherence as the phase relations between related signals is lost.

Clock jitter manifests itself as small oscillations around the threshold of the clock signal, and is mainly constant throughout the full length of the clock signal pulse train.

Clock wander on the other hand is a low frequency modulation of the clock pulses, lengthening and shortening the pulse lengths arbitrarily.

While clock jitter evens out statistically, clock wander is much more significant in value and completely erratic -does not even out for reasonable sampling times.

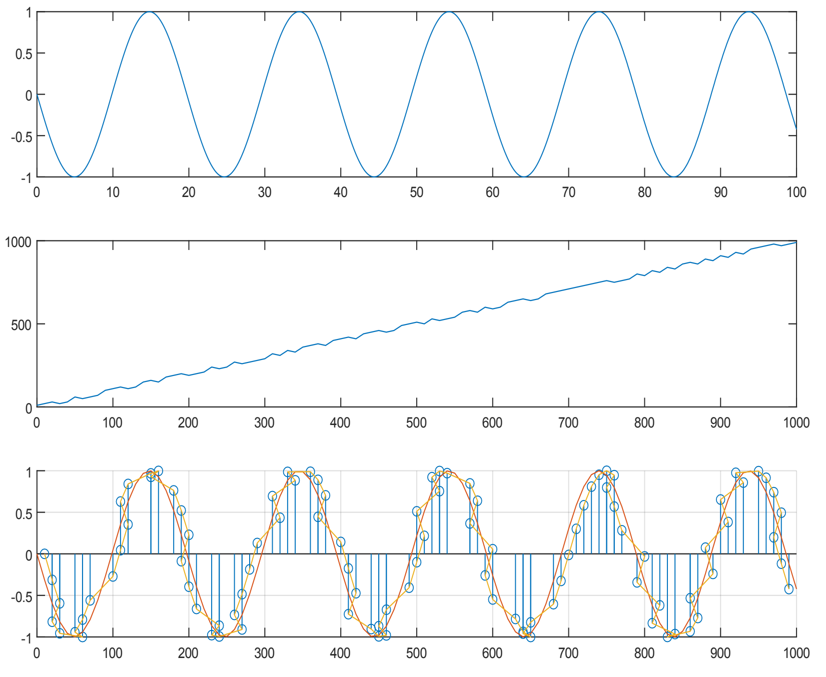


Figure 2. Sampled signal with random (jitter and wander) noise in the acquisition system clock. Panel 1 shows the original signal, panel 2, the jitter affected timing signal compared to the ideal timing signal and reconstituted signal from these samples in panel 3. The phase difference that appears causes synchronization to be lost but also random waveform aberrations.

While the presented case might not seem drastic, consider simultaneous occurring signals or any timestamping operation and take in the fact that the clock jitter is not necessarily the same for different channels of the same data acquisition system and is likely to get worse over time if left uncorrected.

## 2. PRECISE TIMING AS A FUNDAMENT FOR VALID DATA ACQUISITION AND PROCESSING

As we have previously established jitter, both the frame and the individual sample variety, is a source of measurement errors that cannot easily be eliminated after the data was collected. Variations in sampling data rate induce distortions of the waveform and, in the case of frame jitter, per- interval phase errors that lead to loss of coherence between samples of the same signal and between signals.

A model of these distortions is explored in [4] where a method for estimating the standard deviation of frame jitter in data acquired with oscilloscopes and other data collecting equipment. But let us make a simple argument in favor of good timing performance for signal sampling. Let $x[i]$ be a string of correctly sampled values from a signal $f(t)$, and $y[i+\tau(i)]$ a companion string of corresponding sampled values from the same signal, but at erroneous sample intervals $t+\tau(i)$, where $\tau(i)$ is an arbitrary variable sampling delay error caused by jitter. It can be shown that in accordance with equation (6) the original signal can only be recovered by using the $x$ series while the $y$ series reconstitutes a signal with phase and amplitude errors among other distortions. Figure 2 above shows what happens if the sampling system time has small jitter with a measure of random wander.

The third panel graph is the reconstituted signal-showing phase errors and harmonic distortions of an arbitrary nature and rather large values for even small jitter and wander of the sampling system clock.

To some amount, any data collection is subjected to noise in the signal but also temporal quantization noise- that is, jitter. Because of this a number of special data recovery procedures exist- a simplistic approach would suggest that a long enough time interval causes time errors to tend towards a normal distribution, but this is not the case in practice. In fact, the nature of errors in time measurement is such that normal statistic descriptors are of little use as they are cumulative and best described by divergent series.

The question of correctly characterizing time measurement devices led to development of statistic measures such as Allan variance and deviation, and modified Allan variance (MAVAR), total deviation etc.

Data collection and motion control are critical fields for precise timing and synchronization. As we have previously seen in the clock jitter and wander affected data sampling simulation above, the errors are of a random and often accumulating type. It is critical thus to ensure that all data processing systems are in good sync and their on-board oscillators have decent short time precision and reasonable accuracy for the task they are deployed to perform. Jitter, which is constant over time (meaning it cancels out statistically over sufficiently long periods of time), might cause bit errors, but can be filtered in cascaded clock topologies using specialized averaging methods.

Wander can only partly be filtered out in the network nodes and it accumulates in the data transmission network, causing synchronization errors or even a total loss of synchronization. This translates as erroneously decorrelating physically linked phenomena data and inducing large errors in estimates of energies or efforts (forces, moments) or even misrepresenting maximal efforts on sensory elements, e.g., throwing off the overload detection system for the balance etc.

Waveform corruption such as distortion leads to wrong estimates of vibration loads and modes, to wrong estimates of the maximum peak values and to errors in calculating the energy (as area under the curve) where appliable. Dynamic phenomena analysis is rendered impossible.

Clock wanders occur over a period greater than 0.1 s (10 Hz), and vary over time. Therefore, wander measurements must be performed over a longer period of time (e.g., 24 h or longer). Unlike jitter testing, wander testing requires a very stable reference clock (i.e., a Cesium or Rubidium clock) [1], [2], [3].

I feel it is necessary to include a definition of real time systems: they are time-deterministic systems using a special structure that ensures each instruction/command takes a (rigorously fixed) number of processor cycles. While most real time systems are also fast, that is not a requirement. For motion control in loops where necessary data is also collected and sampled, processed and made into commands to the effectors, a real time architecture is required. Also, when correlating data from several sources, a real time system is a must for accurate representation of the investigated phenomena as the phase correlation between signals must be accurate. In practice, at the Trisonic Wind Tunnel there is a somehow more relaxed approach to precise timing and synchronization as the main research focus has been on stable phenomena so far. Recently several tests of highly unstable phenomena and transient regimes became of interest so the new instrumentation has to be capable of accurate synchronization and timing.

## 3. COMMON SYNCHRONIZATION METHODS FOR HETEROGENEOUS AND DISTRIBUTED DATA ACQUISITION SYSTEMS

Heterogeneous data acquisition systems are those comprised of several types of technologies, such as hybrid analog-digital devices, mixed with pure analogic or digital devices, from several vendors and often of different generations. Distributed data acquisition systems are comprised of several partially independent data acquisition sub-systems that are placed in various locations on large scale installations and are kept in sync using some sort of timing and coordination protocol. Industrial solutions for such devices consist of dedicated data buses and protocols such as those used in automation (Ether Cat, SERCOS, Profibus, etc.) but most scientific data acquisition devices use synchronization protocols such as NTP and PTP or even timed precise pulses for synchronous triggering like in 1PPT trigger systems. Simpler systems only use 1PPS or even analog trigger systems. The various industrial systems are complex and use different types of mechanisms for timing data but the principle is essentially the same: either they pass along a shared synchronization and timing device pulse and compensate for delays or use cascades of local clocks that are kept in sync using continuous adjustments to a stable, master clock. This way all data is synchronized and the waveforms are precise. For light testing and demonstration of principles of precise timing reference the author implemented an adaptation of the open-source hardware and software high precision GPS disciplined clock with 1PPS output. The circuit is built around a short time stable, temperature compensated quartz oscillator in the form of a real time clock module with 1 PPS output and a standard GPS receiver (that can accommodate an active antenna for improved indoor reception and reduced frame dropout) that serves as a precise time reference. A generic ATMEL based 8-bit microcontroller and a few lines of code extract the precise timing information from the GPS receiver and set the real time clock, and ensure periodic or on-demand synchronization keeping long term drift and wander at a minimum. With improved short time stability and ensured long-term corrections from the GPS receiver, this type of oscillator is a good starting point for a stable clock reference for data acquisition systems. Figure 3 below shows the breadboarded prototype.

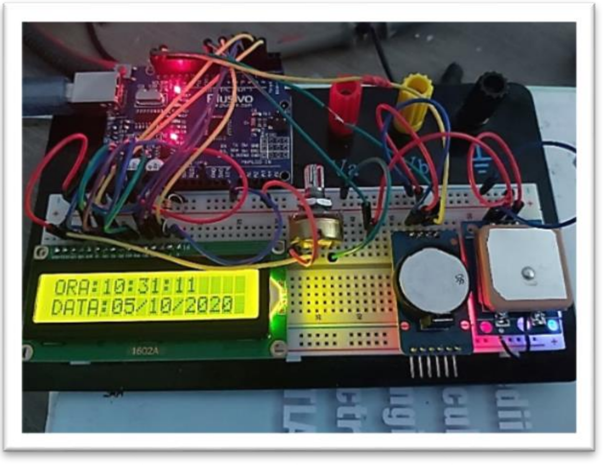


Figure 3. The experimental GPS disciplined oscillator ensures both a long- and short-time stable clock source and 1 PPS (one pulse per second) source for synchronization on data networks. The display and contrast potentiometer are simply for easy monitoring of the clock operation. The 1PPS signal is available on pin 3 of the RTC module, next to the GPS module on the right. The 8-bit generic microcontroller at the top only manages GPS data parsing and real time clock reset tasks, in addition to driving the display.

The prototype shows a good stability and reasonable accuracy; its drift and jitter parameters are currently estimated- we still need a highly stable atomic clock to compare as the device is above or equivalent to the precision class reached by networked time servers (NTP/PTP). Outputs are available in square pulse form, with an adjustable fill factor and frequencies of 1 Hz and 32 KHz, and in text mode format on the serial port on the controller, or via I2C from the real time clock. Also, there is a direct display readout for less demanding synchronization tasks and verification of proper clock operation. Devices similar to this have been in operation both in specialized laboratories and amateur radio transmission shacks for a very long time.

The operation of the prototype is as follows: initially time is read from the GPS receiver and written to the real time clock module's memory. This initial operation, the clock set, is carried out at the initial commissioning. After this step the real time clock will hold the proper time as long as the battery lasts (around 6 years). The second step is operation of the free running real time clock module that also outputs the 1 pulse per second (1PPS) signals with periodic checks against the GPS receiver. The GPS parsing takes around 40 milliseconds on the used hardware controller and is integrated as an offset to the set time whenever necessary. The attainable precision using an intermittent GPS should be better than 100 microseconds [8], as typical with GPS disciplined RTC devices deployed in the field.

Because of the delay caused by the NMEA sentence parsing (the extraction of useful timing data from a general data string containing position data -discarded in this application-, control data, headers, etc provided by the GPS receiver) the 1PPS precision pulse duty is relegated to the real time clock module. This is checked every second for sync errors and whenever they are over a user-specified value, a clock reset operation is carried out by setting the time from the GPS receiver to the RTC. The reset procedure and the parsing procedure are both highly deterministic – they have the same duration, allowing for compensation of delays induced by set/reset and sync operations. As the primary source for jitter and wander of clocks of this design is thermal variations in ambiental temperature, a thermally compensated RTC chip was chosen. Further improvement would see a large thermal mass fitted over all the active components of the RTC as well as a sealed temperature-stabilized container for thermal stability. The prototype was fitted with a passive ceramic-metallic fractal antenna and did not benefit from the increased sensitivity an active antenna would bring. Still because of the high-performance U-Blox Neo 6 receiver it still functioned satisfactorily indoor by the window with the antenna pointed to the open sky.

The prototype ran continuously for two days and at the end of the run no apparent errors were recorded. The 2 microseconds time difference that registered when comparing to a running NTP node on stratum 3(intermittently) with drops to 4 or 16(dropout/local handoff) on a Xeon 1233 V3 based workstation are irrelevant as network latency variability exceeded that value up to 50 microseconds and was not constant over the specified interval. Extracting actual performance data from the prototype is still a matter to be solved satisfactorily by direct comparison with a high precision time standard. Both the hardware and software used are adaptations of the open-source projects released under LGPL license, for example [7] to the available hardware at the moment of drafting this paper. The finished work includes both community and original code and several modifications and adaptations were made to fit the intended purpose, and should be regarded as LGPL compliant- full schematics and code source will be available upon request and filed under LGPL v3. Precisions in similar systems with a local oscillator and on-off GPS to save power were rated at precisions of up to 100 microseconds per day [8].

Using the 1PPS signals several data collecting nodes in a distributed DAQ system can be synchronized far better than the coarse time resolution of one pulse per second might suggest: all data nodes have precise internal clocks capable of subdivisions , but need to be put in sync at the beginning of the data collecting phase, and can use the 1PPS signal as trigger; then, once triggered, they can adjust their variable delay PLLs so that the next pulse finds them in sync; if not, the second on-board high precision timer is used to compute a time difference that is added or subtracted so that the next pulse acquires perfect synchronism.

From the free running tests as well as from literature, the best time adjustment for the GPS disciplined RTC is found to be one using both periodic and aperiodic checks, much like the so-called astronomical regulator clocks used to in the past (Shortt-Synchronome style precision pendulum regulators) [9]. After initial set-up, a second counter is used to check for time deviations on the local working RTC every 30 seconds for example – and adjustments are made as necessary – but also if the deviation is above a certain user-specified threshold, supplemental checks are scheduled, faster than 30 seconds. It might sound puzzling at first but over-correcting a free-running clock is as harmful as under-correcting it- the Allan variance should be as close to linear as possible.

Frequency jumps and unsteady clock rate caused by overcorrecting hurt more the signal stability than a somewhat small running rate error for the base clock. A running rate error leads to signal compression or expansion- harmonic deformation – but sudden frequency jumps lead to a garbled, interrupted or totally broken signal like shown in Figure 2, panel 3.The official clock correction method uses induced jitter to correct for wander, basically trading one type of clock error for smaller quantities of both (jitter and wander)- but clock corrections must take into account the characteristics of each particular oscillator and refrain from overcompensation [10].

From these requirements we deduce the need for quality local oscillators (with low jitter) to ensure minimal short-term signal shape corruption and a well-tempered policy of corrections to keep long term stability within reasonable values.

The project will continue with a long-term stability assessment and building two more prototypes of the same structure, to compare performance and try the improved precision timing technique using “consensus of several independent clocks”. It is also interesting to compare performance of said prototypes with a PTP based PC client using a GPS dongle as a secondary time source. Active GPS antennae will be used to increase sensitivity and improve cold start times for the GPS disciplined clocks.

## 4. CONCLUSIONS

A simple yet effective solution to multiple data source synchronization when using unstructured data is explored. The proposed concept integrates well with older technologies and bridges the gap between raw analog-to-digital converter data and structured data from more modern sources, providing a time stamp to raw data useful when integrating them in a more modern, structured protocol communication capable system. As older systems are phased out there is an interim hybrid state when non timestamped data are required to be precisely aligned to deterministic data (such as those acquired in hard real time systems) and other experiments might require arbitrary synchronization or time stamping across unconnected devices. This cost-effective approach can be implemented and developed to state-of-the-art standards and allows using different generations of test equipment in the same experimental facility.

## ACKNOWLEDGEMENTS

The author thanks the organizing committee of *The International Conference of Aerospace Sciences "AEROSPATIAL 2020", Virtual Conference, 15-16 October 2020, Bucharest, Romania,* conference where this work was first presented in a simpler form.

## REFERENCES

[1] A. N. Kalashnikov, R. E. Challis, M. E. Unwin and A. K. Holmes, Effects of frame jitter in data acquisition systems, in *IEEE Transactions on Instrumentation and Measurement*, vol. **54**, no. 6, pp. 2177-2183, Dec. 2005, doi: 10.1109/TIM.2005.858570.

[2] T. Dabóczi, Uncertainty of signal reconstruction in the case of jittery and noisy measurements, *IEEE Transactions on Instrumentation and Measurement*, vol. **47**, no. 5, pp. 1062–1066, Oct. 1998.

[3] T. M. Souders, D. R. Flach, C. Hagwood, and G. L. Yang, The effects of timing jitter in sampling systems, *IEEE Transactions on Instrumentation and Measurement*, vol. **39**, no. 1, pp. 80–85, Feb. 1990.

[4] A. N. Kalashnikov, R. E. Challis, M. E. Unwin, and A. K. Holmes, *Quantification of frame jitter in data acquisition systems*, in Proc. 2003 IEEE Int. Symp. Intelligent Signal Proc., ch., Budapest, Sep. 2003, pp. 15–20.

[5] R. N. Bracewell, *The Fourier Transform and Its Applications (revised ed.),* 1986, McGraw-Hill, 1st ed. 1965, 2nd ed. 1978.

[6] M. Rahman, *Applications of Fourier Transforms to Generalized Functions*, WIT Press Southampton, Boston, ISBN 978-1-84564-564-9, 2011.

[7] * * * *A cheap and accurate clock based on GPS - Arduino Project Hub*, retrieved 27-11-2020, an open-source project released under LGPL.

[8] R. Tian, J. Zhang, S. Zhang, L. Wang, H. Yang, Y. Chen, Y. Jiang, J. Lin and L. Zhang, A High-Precision Energy-Efficient GPS Time-Sync Method for High-Density Seismic Surveys Applied Sciences, *Applied Sciences*, An Open Access Journal from MDPI, 2020.

[9] A. M. Panait, *Shortt-Synchronome Time synchronization in distributed data collection systems- an old solution to a new problem,* AEROSPATIAL 2020, Virtual Conference, Online, 15th of October 2020.

[10] R. H. Miles, *Synchronome - Masters of Electrical Timekeeping*, London: AHS. ISBN 978-0901180551, 2019.