Clock Composition by Wiener Filtering Illustrated on Two Atomic Clocks

Marek Peca*, Vojtěch Michálek, Michael Vacek
Serenum, a. s., Beranových 130, Praha – Letňany, Czech Republic
České vysoké učení technické v Praze, Břehová 7, 115 19 Praha 1, Czech Republic
*e-mail: peca@vzlu.cz

***Abstract*— Estimation instead of feedback loops is recommended to obtain a composite clock. Wiener filtering approach to clock ensembling is introduced and demonstrated on the simplest case of two clocks. Design procedure dealing with clock system non-stationarity, non-observability and numerical issues, is given. Impact of causality to unexpected performance degradation is discussed.**

I. Introduction

Composite clock, or clock ensembling, i. e. calculation of best time estimate given readings from multiple clocks, is a must for state-of-art timekeeping. There are two distinct tools for clock ensembling: (i) feedback; (ii) estimation. Using (i), a phase- or frequency-locked loop or loops (PLL, FLL) are formed, containing clock and a controller. The controller corrects clock's time (by tuning, modulation, phase stepping, etc.), and the corrected signal is fed back into the controller. On the other hand, in case of (ii) an estimator senses clocks' reading without any modification, and produces estimate of clocks' state, applicable as a correction to time reading. The corrected signal is not fed back into the estimator.

Realizations of (i) are e. g. Network Time Protocol (NTP), and Atomic Clock Ensemble in Space (ACES [1]). Suppose the controller (i) is constrained to be linear (what is the common case). Then, following separation principle [2], even the optimal controller (minimum variance controller) will give worse or same performance, as an optimal linear estimator (ii); the limit case of the same performance requires zero control noise. Therefore, we claim that use of (i) is justified only when implementation of estimator is not feasible (e. g. in case of specific analogue circuitry). Otherwise, including ACES and NTP in our opinion, the estimator (ii) is the right way to choose.

II. Clock modelling

The essential prerequisite for design of ensembling is a clock model, describing statistical properties of phase evolution over time $x(t)$. A clock is described as a linear stochastic system, defined by phase spectrum $\mathcal{S}_{xx}(f)$. The system is marginally stable: it contains one or more integrators in cascade, corresponding to $f^{-n}, n = 2, 4, \ldots$ terms in $\mathcal{S}_{xx}(f)$. In addition, it may contain stable modes as well. Within estimation approach to ensembling, clocks are not disciplined, the system is purely stochastic, no deterministic input (tuning).

A discrete-time processing is assumed further. Estimates are calculated digitally from sequence of measurements. For convenience, these shall be acquired at equidistant time instants $t_k, t_k - t_{k-1} = T_s = const.$. This can be fulfilled only approximately, because T_s is perturbed by clock noise as well. Although considered by [3], we follow the conclusion of [4], that the effect is negligible for the purpose of ensembling.

Confinement to a class of linear discrete-time systems, containing stable and marginally stable modes, is sufficient for estimator design procedure (Sec. III). Assumption of clock process linearity implies that properly designed linear estimator is optimal in sense of mean squared error (MSE). It also implies, that spectrum $\mathcal{S}_{xx}(f)$ is sufficient description of the process – possibly different clocks with equal $\mathcal{S}_{xx}(f)$ are indistinguishable, regardless of their internal structure (Spectral factorization theorem [2], [5]).

Assumption of linearity works well for f^{-2}, f^{-4}, f^{-6} spectrum terms, which indeed origin from cascaded integration of error. A more peculiar component of the spectrum is $1/f$-noise (flicker), constituting f^{-3}, f^{-5} terms [6]. Works on deterministic chaos suggest an inherently non-linear behavior as a cause of $1/f$-noise [7]. The $1/f$-noise can not be generated by linear system of finite order. However, for a given frequency band of interest and required fidelity, the spectrum may be approximated by a linear system of some finite order. The interesting question follows, whether a linear estimator designed for such an approximate system may approach the optimal estimator even for a process, containing non-linear (chaotic) $1/f$-noise. The answer is not known to us, so we follow [3], approximating $1/f$-noise terms by finite-order, linear, discrete-time model.

A single-input, single-output (SISO) discrete-time linear system can be described by its transfer function $G(z)$: $x(t) = G(z)u(t)$ (input signal u transformed to output y; z is a forward shift operator). In our case of purely stochastic system (no control input) $u(t)$ is a unit variance white noise. For finite-order systems $G(z) = B(z)/A(z)$ where B, A are polynomials. A spectrum of x ($S_{xx}(z)$ or $\mathcal{S}_{xx}(f)$) is given by:

$$S_{xx}(z) = G^*(z)G(z)S_{uu} = \frac{B^*(z)B(z)}{A^*(z)A(z)}S_{uu}, \qquad (1)$$

$$\mathcal{S}_{xx}(f) = S_{xx}(e^{j\theta}), \quad \theta = 2\pi f T_s,$$

where $G^*(z) = \overline{G(1/z)}$. If x is real-valued, $S_{xx} = S_{xx}^*$. Given S_{xx} can be always factored as product of G, G^* (1) so that roots of B, A lie inside the unit disc. The step is called spectral factorization, denoted $G(z) = \mathrm{spf}(S_{xx}(z))$, producing stable, minimum-phase model G [2], [5]. $\mathrm{spf}(\cdot)$ allows to create clock model out of a given phase spectrum. A state-space clock model $\mathbf{x}(t+1) = \mathbf{\Phi}\mathbf{x}(t) + \mathbf{u}(t)$ [4], [8], [3], [9], [10] can be converted to SISO $G(z)$ by standard means [2].

The model of single clock has been given. Ensemble is a set of N clocks (Fig. 1a). Using any meaningful physical means of time signal processing (counters, phase comparators, mixers, etc.), only time differences between the clocks may be measured. There is no clue about an "absolute" time offset. With any number of mutual measurements, we end up with only $N-1$ degrees of freedom in data: measurement matrix is singular.

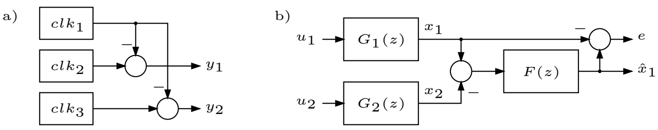

Fig. 1. (a) Clock ensemble (b) Composition of two clocks

Two important consequences follow, whose conjunction makes the estimation task non-trivial one: (i) time offset (error) of each individual clock grows without bounds; (ii) clock ensemble system is not completely observable. Therefore, any possible time estimator produces time estimate whose error grows without bounds as well. The goal is to achieve the lowest possible error within finite horizon. It is probably this specialty, why the topic of clock ensembling is still in active research [9], [10].

III. Estimator design

The most general form of MSE optimal linear estimator is Kalman filter (KF [2]). The mentioned clock ensemble's non-stationarity and non-observability cause difficulties to practical KF computation due to unbounded covariance growth. This problem gave rise to KF variants with specific covariance treatment [4], [8], [3], [10]. We have chosen another way, leading to a simple, closed-form estimator: Wiener filter (WF [5], [2]). The only limitation of WF for clock applications is that time-invariant processes are assumed, i. e. clock spectra supposed to be constant during estimation. In the simplest case of only two clocks [1], the WF becomes SISO, and the expressions are very simple.

The ensemble of $N = 2$ clocks $G_{1,2}(z)$ thus produces only one $(N-1)$ measurement y, Fig. 1b. $G_{1,2}$ are assumed SISO, i. e. $\mathrm{spf}(\cdot)$ performed if necessary. The time difference y is fed into estimator $F(z)$, which is designed to estimate output of one of the two clocks, $\hat{x}_1$. The $\hat{x}_1$ samples constitute corrections of x_1 signal, reducing effectively an uncertainty of composite time to that of residual signal e. The task is symmetrical, after exchanging $G_1 \leftrightarrow G_2$, the estimator becomes $F'(z) = 1 - F(z)$. Optional measurement noise may be incorporated into G_2 model.

The design of linear, MSE optimal estimator following Wiener formalism is based on two spectra: spectrum S_{yy} of the measured signal y, and cross-spectrum between y and signal to be estimated x, designated S_{xy}. In our case $y = x_1 + x_2$ and $x_{1,2}$ are uncorrelated ($x_1 \perp x_2$) therefore simply $S_{xy} = S_{xx}$.

There are three different variants of resulting estimators: non-causal $F_{nc}(z)$, causal $F(z)$ and finite-lag $F_T(z)$ WF. $F_{nc}(z)$ is best (lowest MSE), but requires to process future samples $y(t+1) \ldots y(+\infty)$. It can not be used in real-time, only offline in batch processing (smoothing). $F(z)$ is designed to deliver estimates without any lag, within the measurement cycle, possibly in real-time. Because of the lack of future y development information, its MSE is worse. A gap between $F_{nc}(z)$ and $F(z)$ is filled with $F_T(z)$, allowing to trade performance vs. filter lag T (number of future samples to wait for). The non-causal solution is simple [5], [2]: $F_{nc} = S_{xy}/S_{yy}$; in our specific case:

$$F_{nc} = \frac{S_{xx}}{S_{yy}} = \frac{S_{11}}{S_{11} + S_{22}},$$

where $S_{kk} = G_k^* G_k$ are clock spectra; the residual error is:

$$e = x_1 - \hat{x}_1 = \frac{1}{w_1 + w_2}(w_1 x_1 + w_2 x_2), w_k = \frac{1}{S_{kk}}.$$

Residual spectrum is easily plotted, or even sketched:

$$S_{ee} = \frac{S_{11} S_{22}}{S_{11} + S_{22}}$$

.

Considering $\mathcal{S}_{xx}(f) = \sigma_x^2(f)\mathrm{d}f$, we see that estimator averages input signals $x_{1,2}$ weighted by inverse of their respective variances at given frequency. If both clocks have a spectrum of the same shape $S_{11} \propto S_{22}$, estimator reduces to mere static weighted average $F_{nc}(z) = const.$, and it is also the only case when $F_{nc}(z) = F(z)$ is causal.

If both processes $x_{1,2}$ (or the output y), are filtered by common transfer function $G_1'(z) = G_c(z)G_1(z)$, $G_2'(z) = G_c(z)G_2(z)$, then $F_{nc}(z)$ remains unchanged. Thanks to this property, $F_{nc}(z)$ is equally optimal for estimation of time (phase), as for estimation of frequency.

$F(z)$ is not allowed to weight future data so its impulse response must be zero in negative time, $h(t < 0) = 0$. Operation to truncate $H(z)$ to its causal part is denoted $[H(z)]_+$. A naïve approach to $F(z)$ might be to take $[F_{nc}(z)]_+$. This resembles a common practice of block data processing: samples outside of a dataset are expected to be zero. Despite of that, $[F_{nc}(z)]_+$ is not the MSE optimal $F(z)$. The right solution is derived with help of a notion of whitening filter $W(z)$ [5]:

$$W(z) = \mathrm{spf}(1/S_{yy}), \qquad F(z) = [S_{xy}W^*]_+ W.$$

Compare $F_{nc}(z) = (S_{xy}W^*)W$ – only a part of $F_{nc}(z)$ underwent the causal truncation. $F(z)$ is no longer invariant to multiplication by common factor G_c. Therefore, MSE optimum for time generally differs from MSE optimum for frequency in case of causal estimator. $F_T(z)$ is similar to $F(z)$, only the $(S_{xy}W^*)$ is allowed to look T samples into the future: $F_T(z) = z^T[z^{-T}S_{xy}W^*]_+W$. The properties are similar to that of $F(z)$, performance compares as follows: var $e_{causal} \geq$ var $e_T \geq$ var e_{nc}.

Design procedure begins with clock models in form of polynomial fractions $G_1(z) = B_1(z)/A(z)$, $G_2(z) = B_2(z)/A(z)$, common denominator $A(z)$ is assumed. Models may be safely expanded to common $A(z)$ if required. Spectrum of measured signal is:

$$S_{yy} = (U_1^*G_1^* + U_2^*G_2^*)(G_1U_1 + G_2U_2)$$

Since $u_1 \perp u_2$, var u_1 = var $u_2 = 1$, $U_1^*U_2 = 0, U_1^*U_1 = U_2^*U_2 = 1$:

$$S_{yy} = G_1^*G_1 + G_2^*G_2 = \frac{B_1^*B_1 + B_2^*B_2}{A^*A} = \frac{C^*C}{A^*A}$$
$$\text{where} \quad C = \mathrm{spf}(B_1^*B_1 + B_2^*B_2). \tag{2}$$

Note $C(z)$ is obtained by $\mathrm{spf}(\cdot)$ (2). Cross-spectrum of $x = x_1 \to y$ is:

$$S_{xy} = (U_1^*G_1^*)(G_1U_1 + G_2U_2) = G_1^*G_1 = \frac{B_1^*B_1}{A^*A}$$

.

Giving the non-causal WF:

$$F_{nc} = \frac{S_{xy}}{S_{yy}} = \frac{B_1^*B_1}{C^*C}.$$

Causal WF follows:

$$W = A/C, \text{ so that } S_{yy} = 1/(W^*W)$$
$$F = [S_{xy}W^*]_+W = \left[\frac{B_1^*B_1}{AC^*}\right]_+ \frac{A}{C} = \left[\frac{D_+}{A} + \frac{D_-}{C^*}\right]_+ \frac{A}{C}$$
$$B_1^*B_1 = D_+C^* + D_-A \tag{3}$$
$$F(z) = \frac{D_+(z)}{C(z)}. \tag{4}$$

The polynomials $D_{+,-}$ result from (3). Their orders should be constrained so that D_+/A is causal (and possibly containing absolute term $z^0 \equiv h(0)$), while D_-/C^* is strictly anti-causal, non-containing the absolute term. Such a constraint assures unique solution to (3) [11].

IV. Specific problems

The central problem of clock ensembling is the non-stationarity and non-observability of the system (Sec. II). The non-stationarity manifests itself as a marginally-stable $(z-1)^m$ factor: $A(z) = (z-1)^m\tilde{A}(z)$, where $\tilde{A}(z)$ are stable factors (possibly $\tilde{A}(z) = 1$ for pure integrator models). Some of the WF formulations disallow marginally stable factors at all; others [11] do allow them, but they require at least the residual error e variance to be bounded. In our case, this condition is not satisfied. To overcome problems caused by $(z-1)^m$ ($\mathcal{S}_{\dots}(f \to 0) \to \infty$), we employ following notional alteration $A(z) = (z-(1-\epsilon))^m\tilde{A}(z)$.

Integrators were substituted by 1^{st}-order low-pass filters with cutoff $f_c \to 0$ smaller than any interesting frequency in the system. An important feature of this workaround is that the infinitesimally displaced factor $(z-(1-\epsilon))$ is interchangeable with pure $(z-1)$ in actual computation: $A(z)$ does not enter (2) at all, and there are no roots close to $(z-1)$ in (3) except $A(z)$ itself. The only purpose of said alteration is to determine that $A(z)$ is a stable, or causal (4) polynomial, and $A^*(z)$ is an unstable, anti-causal one. Observe also that $A(z)$ cancels out of both $F_{nc}(z)$ as well as $F(z)$ result.

Another design difficulty is due to a huge range of interesting frequencies to be modelled in clock ensembles. E.g., in [1] two-clock ensemble, a key area of interest lies around band from 5×10^{-6} Hz to 1 Hz. Suppose system sampling frequency is $f_s = 10$ Hz. A filter requested to emphasize or suppress signal in given band needs to contain poles located near $z = 0.54$, $z = 0.999997$. This makes (2) hard or impossible to solve by means of root-finding in ordinary double-precision floating point arithmetics (64b FP). Therefore, we have switched to arbitrary-precision arithmetic software (Maple 9.5). The $\mathrm{spf}(\cdot)$ (2) has been solved by addition, multiplication, root-finding, and discarding all $|z| > 1$ roots.

The second step consists in solving (3), leading to a system of linear algebraic equations. We have not noticed numerical difficulties here using 64b FP, but anyway we continued to solve the linear system in arbitrary-precision domain as well. For the examples described below, a precision of hundred(s) of decimal digits always yielded plausible results.

The design procedure yields an estimator as a polynomial fraction such as $F(z) = D_+(z)/C(z)$ in the causal variant. It is an infinite impulse response (IIR) filter, whose modes correspond to roots of $C(z)$. The $C(z)$ should be stable by definition of $\mathrm{spf}(\cdot)$, unless (2) fails to compute due to insufficient arithmetic precision.

Filter coefficients come from (2,3) in overly large precision, unjustified for practical implementation. Implementation of the filter using chosen word length and calculation structure, stability and performance might be degraded by means of: (i) signal round-off error and its propagation; (ii) filter coefficient displacement. Respective countermeasures belong to field of DSP expertise [12]. Good tool to asses implementation performance shift due to (ii) is residual spectrum S_{ee}. In our examples, we have used 64b FP and direct form IIR structure.

It should be stressed that the stability of whole clock composition system lies in the estimator. Therefore, if the IIR filter implementation is stable on finite wordlength arithmetic level, the whole system is guaranteed to be stable. This is a remarkable difference from the ensembling systems relying on feedback loops (PLL, FLL, etc.), where an improper matching of system model to physical clocks or tight stability margin may cause instability.

V. Examples

First example is artificial; both clocks follow the same model $x_k(t) = (r_{k,1}/(z-1)^2)u_{k,1} + r_{k,2}u_{k,2}$, but different parameters. S_{11}, S_{22}, and S_{ee} for both $F(z)$, $F_{nc}(z)$ is shown in Fig. 2. $F(z)$ is apparently worse than $F_{nc}(z)$, interestingly, in small region it is worse than any of the two input clocks. Still, it remains optimal by means of S_{ee} integral over f.

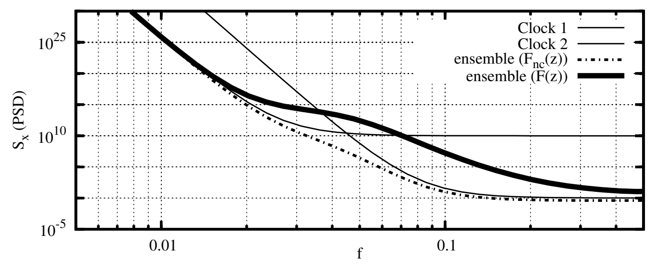

Fig. 2. Comparison of non-causal vs. causal WF for artificial clocks

Second, WF has been applied to two-clock (Cs clock “PHARAO”, H-maser “SHM”) ensemble model, as published [1]. $1/f$-noise of SHM has been approximated by 3^{rd}-order model, Fig. 3. While $F_{nc}(z)$ acts as expected, $F(z)$ is a real surprise: the time (phase) MSE-optimal causal estimator almost completely discards the SHM reading, weighting it up to $|F(z)|_{max} = 0.063$. S_{ee} is nearly identical to spectrum of PHARAO.

Is it correct? For MSE optimum in time ($\equiv$ phase), yes. By intuition, this probably is not the desired solution. We suppose, that the best way to get the most of the two clocks is to employ $F_T(z)$, in the graph plotted for chosen lag $T = 10^4\,\mathrm{s}$.

Recall, that any (linear) feedback system, as the present PLL&FLL solution [1], must be same or practically worse than even $F(z)$ by means of MSE. Performance of $F(z)$ and $F_{T=10^4\,\mathrm{s}}(z)$ is compared to PLL&FLL in Allan variance graph, Fig. 4. It is worth mentioning that PLL&FLL implementation is much more complex than that of IIR WF in this case.

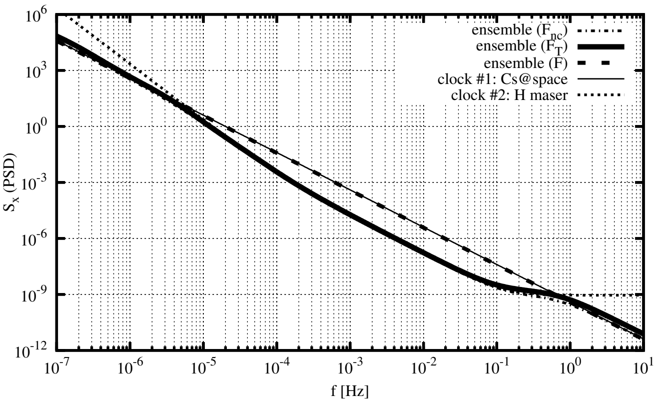

Fig. 3. Clock and residual phase spectra of ACES clocks and WF

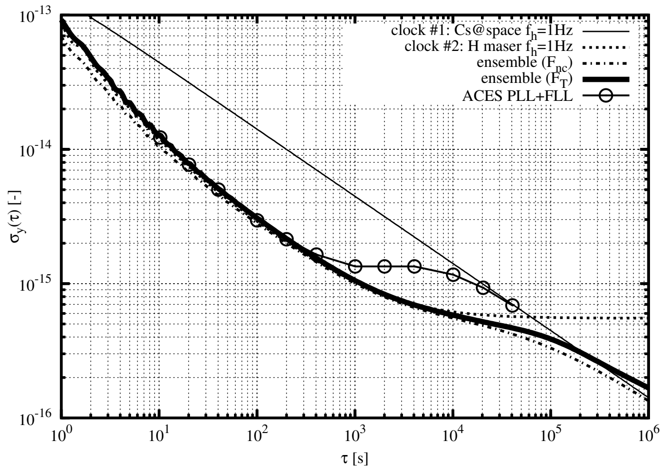

Fig. 4. Allan deviation of ACES clocks and WF residuals

VI. Conclusion

Use of estimation approach to clock composition instead of feedback loops is encouraged, wherever estimator implementation is possible. Design of WF for two-clock ensemble has been provided, dealing with non-observability of the system. Practical example of the ACES project model shows, how WF outperforms current PLL&FLL-based solution in performance as well as simplicity. Besides this, it shows how significant may be an advantage of finite-lag over causal WF.

References

[1] J. K. Dam, W. Schäfer *et al.*, “Simulation of Servo Loops in Atomic Clock Ensemble in Space (ACES),” in *EFTF 2010 – 24th European Time and Frequency Forum*, 2010.

[2] K. J. Åström and B. Wittenmark, *Computer Control Systems: Theory and Design*, 2nd ed. Prentice Hall, 1990.

[3] J. A. Davis, C. A. Greenhall, and P. W. Stacey, “A Kalman filter clock algorithm for use in the presence of flicker frequency modulation noise,” *Metrologia*, vol. 42, no. 1, pp. 1–10, 2005.

[4] K. R. Brown, “The Theory of the GPS Composite Clock,” in *Proceedings of ION GPS-91*, Albuquerque, New Mexico, September 1991, pp. 223–241.

[5] E. W. Kamen and J. K. Su, *Introduction to Optimal Estimation*. Springer London, 1999.

[6] J. A. Barnes, A. R. Chi *et al.*, “Characterization of Frequency Stability,” *IEEE Transactions on Instrumentation and Measurement*, vol. IM-20, pp. 105–120, 1971.

[7] H. G. Schuster and W. Just, *Deterministic Chaos: An Introduction*, 4th ed. Wiley-VCH Verlag GmbH, 2005.

[8] C. A. Greenhall, “Forming stable timescales from the Jones-Tryon Kalman filter,” *Metrologia*, vol. 40, no. 3, pp. 335–341, 2003.

[9] L. Galleani and P. Tavella, “Time and the Kalman Filter,” *IEEE Control Systems*, vol. 30, no. 2, pp. 44–65, April 2010.

[10] C. A. Greenhall, “A Review of Reduced Kalman Filters for Clock Ensembles,” *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 59, no. 3, pp. 491–496, March 2012.

[11] A. Ahlén and M. Sternad, *Polynomial Methods in Optimal Control and Filtering*, K. Hunt, Ed. Institution of Engineering and Technology, 1993, Chapter 5: Optimal Filtering Problems.

[12] B. W. Bomar, *The digital signal processing handbook*. CRC Press, 1998, ch. Finite wordlength effects.