Effect of loop delay on phase margin of first-order and second-order control loops

Citation for published version (APA):

DOI:
10.1109/TCSII.2005.852003

Document status and date:
Published: 01/01/2005

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Effect of Loop Delay on Phase Margin of First-Order and Second-Order Control Loops

Jan W. M. Bergmans, Senior Member, IEEE

I. INTRODUCTION

Data receivers for digital transmission and storage systems normally contain various control loops (e.g., automatic gain control, adaptive dc compensation, adaptive equalization, and timing recovery) that jointly act to permit reliable data recovery in spite of varying or uncertain system conditions [1], [2]. Data rates and computational requirements in these systems tend to outpace Moore’s law [3], [4]. This spurs increasing use of techniques such as pipelining and parallelization. As a result, loop delays increase and can become a limiting factor, especially during acquisition.

Control loops in data receivers are often decision-directed, i.e., control information is derived with the aid of the bit decisions. Bit detectors necessarily become increasingly powerful, and, as a consequence, their detection delay increases, thus further increasing loop delay. Especially during acquisition, the compound loop delay can become prohibitively large. A typical approach to mitigate this problem involves the use of auxiliary detectors that produce tentative decisions with minimum delay [5], [6].

Against the above background, it is increasingly important to understand how loop delay affects the properties of the loop. Previous studies have been directed at identifying the edge of the stability region of first-order and second-order discrete-time loops in the presence of a loop delay [7], [8]. In practice, it is desirable to operate loops well within the stability region. In this respect, the so-called phase margin is often used as a measure of the degree of loop stability, and it is desirable to be able to dimension the loop for a prescribed phase margin. This paper analyzes the impact of loop delay on the phase margin of first- and second-order control loops of both the discrete-time and the continuous-time variety, with the remainder of the paper organized as follows. Section II analyzes the first-order loop, which is representative, for example, of automatic gain control (AGC) loops, dc control loops, and adaptive equalizers. Section III focuses on the second-order discrete-time high-gain loop. Since second-order control loops are mainly found in phase-locked loops (PLLs), this section is cast in PLL terms. This also applies to Section IV, which focuses on the second-order continuous-time high-gain loop. Behavior of this loop can approximate that of its discrete-time counterpart, yet analytical results are comparatively simple and hence insightful. For both loops, an adequate phase margin is required to limit jitter. Section V establishes rules of thumb for accomplishing a prescribed phase margin and draws conclusions. Detailed analysis for the second-order discrete-time and continuous-time loops is relegated to Appendices I and II.

II. FIRST-ORDER CONTROL LOOP

We first consider the discrete-time first-order loop (Fig. 1). The ideal value of the control parameter is denoted \( x_k \), the actual value is denoted \( y_k \), and the error \( x_k - y_k \) is denoted \( e_k \) (the subscript \( k \) denotes the time index expressed in sampling intervals \( T \)). The error \( e_k \) is delayed by \( M \) sampling intervals \( T \) and scaled by a compound loop gain \( K_f \) which determines loop bandwidth and tracking speed. The loop is closed via a first-order ideal integrator (the symbol \( z^{-1} \) denotes a delay of one sampling interval \( T \)). The model of Fig. 1 is illustrative for, e.g., AGC and dc compensation loops. A typical gain \( K_f \) for such loops is \( K_f \approx 0.1 \) during acquisition. After acquisition, a substantially lower value (e.g., \( K_f \approx 0.01 \)) is normally used for parameter tracking. In the absence of a delay \( M \), the loop has a first-order exponential impulse response with a time constant \( \tau = T/K_f \) (expressed in seconds). We can think of \( \tau/T \) as the normalized time constant, i.e., the time constant expressed in symbol intervals \( T \). Clearly, \( \tau/T = 1/K_f \).

The loop has closed-loop transfer function

\[
G(z) = \frac{E(z)}{X(z)} = \frac{1}{1 + H(z)}
\]

\[Manuscript received October 9, 2003; revised April 3, 2004. This work was supported in part by the European Union under the ‘TwoDOS’ IST Project IST-2001-34168. This paper was recommended by Associate Editor X. Yu.

The author is with Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands (e-mail: J.W.M.Bergmans@tue.nl).

Digital Object Identifier 10.1109/TCSII.2005.852003
where $H(z) = z^{-M}K_l I(z)$ is the open-loop transfer function and $I(z) = z^{-1} / (1 - z^{-1})$ is the transfer function of the discrete-time integrator in Fig. 1 (we use capitals to denote the $z$ transforms of the corresponding lower case sequences). In frequency-domain notation, we may write

$$H(e^{j2\pi \Omega}) = \frac{K_l}{2 \sin \pi \Omega} e^{-j2\pi \Omega(2M+1)} = \frac{K_l}{2 \sin \pi \Omega} e^{-j\phi(\Omega)}$$

with

$$\phi(\Omega) = -\pi \left[ \frac{1}{2} + 2(M + 0.5)\Omega \right].$$

Here, $\Omega$ is a normalized measure of frequency, with $\Omega = 1$ corresponding to the sampling rate $1/T$. The phase margin PM is determined at the unity-gain frequency of $H(e^{j\omega})$, i.e., at the frequency $\Omega = \Omega_0$ for which $|H(e^{j\omega})| = 1$. The system will be unstable if the phase $\phi(\Omega)$ of $H(\Omega)$ exceeds $-\pi$ rad at $\Omega = \Omega_0$. By definition [9, Sec. 6.4], PM is the margin that is left with respect to this stability limit, i.e.,

$$PM = \phi(\Omega) = -\pi + \phi(\Omega_0).$$

In practical systems one typically requires a phase margin of around $45^\circ$ to $60^\circ$, i.e., around $\pi/4$ to $\pi/3$ radians.

For the loop at hand we evidently have $\Omega_0 = (1/\pi) \arcsin [K_l/2]$. Correspondingly

$$PM = \pi + \phi(\Omega_0) = \frac{\pi}{2} - 2(M + 0.5) \arcsin \left( \frac{K_l}{2} \right).$$

For loop gains of practical interest (i.e., for $K_l \ll 2$) we have $\arcsin[K_l/2] \simeq K_l/2$, whence $\Omega_0 \simeq K_l/(2\pi)$ and

$$PM \simeq \frac{\pi}{2} - (M + 0.5)K_l.$$

Evidently PM decreases as $M + 0.5$ increases. The contribution $0.5$ can be regarded as the effective delay introduced by the integrator in Fig. 1, and we can think of $M + 0.5$ as the compound effective loop delay.

The edge of the stability region is demarcated by PM = 0. Here $K_l \simeq \pi/[2(M + 0.5)]$. Phase margins of $\pi/4$ ($45^\circ$) and $\pi/3$ ($60^\circ$) are obtained for $K_l \simeq \pi/[4(M + 0.5)]$ and $K_l \simeq \pi/[6(M + 0.5)]$, respectively. In terms of the normalized time constant $\tau / T = 1/K_l$, we have

$$M + 0.5 \simeq \frac{\tau}{T} \left[ \frac{\pi}{2} - PM \right].$$

This equation reveals the largest loop delay that is permissible to achieve a prescribed phase margin. In particular for a phase margin of $45^\circ$ ($\pi/4$), $M + 0.5$ should be no larger than $\pi/4$ times the normalized time constant $\tau / T$, versus $\pi/6 \simeq 0.5$ times for a phase margin of $60^\circ$ ($\pi/3$).

Continuous-Time Loop: The model of this loop is that of Fig. 1, but with all discrete-time quantities replaced by their continuous-time counterparts. Specifically, the discrete-time integrator is replaced by a continuous-time integrator (with transfer function $1/(j\omega T)$ where $\omega$ denotes the angular frequency), and the discrete loop delay of $M$ sampling intervals is replaced by a continuous delay $\lambda = (M + 0.5)T$. Upon retracing the above steps for this continuous-time model, one readily verifies that all approximate equality signs above become exact equalities. This reflects the fact that the continuous-time and discrete-time loops behave essentially identically for $K_l \ll 2$ [2, Ch. 11]. Hence, the above results carry over directly to the continuous-time case.

### III. SECOND-ORDER DISCRETE-TIME PHASE-LOCKED LOOP

This loop is of the high-gain second-order type and has the discrete-time model of Fig. 2. The input phase $\psi_k^*$ is tracked by a voltage-controlled oscillator (VCO) that is modeled as an ideal integrator with output phase $\psi_k$. The phase error $\Delta_k^* = \psi_k^* - \psi_k$ (normalized in sampling intervals $T$) serves as the VCO input $\eta_k$ after being delayed across $M$ sampling intervals and filtered by the loop filter. This filter has a proportional and an integrating path with total open-loop gains $K_p^*$ and $K_f^*$, respectively. Any gain of the phase detector and/or VCO is accommodated in $K_p^*$ and $K_f^*$. These two parameters together determine the normalized natural frequency $\omega_n T$ and damping factor $\zeta$ of the PLL according to [2, Ch. 11]

$$\omega_n T = \sqrt{K_f^*} \quad \text{and} \quad \zeta = \frac{K_p^*}{2\sqrt{K_f^*}}.$$  \hspace{1cm} (1)

A typical loop uses $\omega_n T \simeq 0.1$ during acquisition and $\omega_n T \simeq 0.01$ during tracking. In both cases, the damping factor $\zeta$ is in the order of unity.

It is worth mentioning that the notions of natural frequency and damping factor pertain to a second-order system and are, hence, strictly speaking, only applicable in the absence of a delay $M$. For the sake of consistency, we will nevertheless use them here, defined as in (1), for any value of $M$.

The phase margin of the loop is derived in Appendix I. For various values of $M$, Fig. 3 portrays the combinations of $\omega_n T$ and $\zeta$ that yield a phase margin of $60^\circ$. Similar graphs are shown in Fig. 4 for phase margins of $30^\circ$ and $0^\circ$. The latter margin corresponds to the edge of the stability region, which was derived earlier in [7].

Irrespective of $M$, for phase margins of practical interest, the curves all have a maximum for damping factors in the order of unity. They level off only slowly as $\zeta$ increases beyond this optimum, yet decline rapidly as $\zeta$ decreases. This suggests that a proper engineering choice for $\zeta$ would be a value somewhat above unity.
Fig. 3. Normalized natural frequency $\omega_n T$ versus damping factor $\zeta$ for second-order discrete-time PLL with a phase margin of 60°. The loop delay $M$ is used as a parameter.

Fig. 4. As Fig. 3, but for phase margins of 30° and 0°.

IV. SECOND-ORDER CONTINUOUS-TIME PHASE-LOCKED LOOP

The phase margin of the continuous-time PLL is derived in Appendix II and has a considerably simpler analytical form than that of the discrete-time PLL. Specifically,

$$\text{PM} = \arctan\left[2\zeta F(\zeta) - F(\zeta)\omega_n \lambda\right]$$

(2)

where $F(\zeta) = \sqrt{2\zeta^2 + \sqrt{1 + 4\zeta^2}}$ is a monotonically increasing function of the damping factor $\zeta$ and $\lambda$ is the effective loop delay in seconds. In terms of the discrete-time loop of

Fig. 5. Normalized natural frequency $\omega_n T$ versus damping factor $\zeta$ for second-order continuous-time PLL with phase margin of 60°. The loop delay $\lambda$ amounts to $(M + 0.5)T$.

Fig. 6. Same as Fig. 5, but for phase margins of 30° and 0°.

Fig. 2, we may equate $\lambda$ with $(M + 0.5)T$ where the contribution $0.5T$ accounts for the effective delay of the discrete-time integrator that models the VCO.

In the absence of a loop delay (i.e., for $\lambda = 0$), PM is fully determined by $\zeta$ and does not depend on $\omega_n$. The presence of a loop delay causes PM to decrease in linear proportion to the normalized loop delay $\omega_n \lambda = (M + 0.5)\omega_n T$.

Figs. 5 and 6 are the counterparts of Figs. 3 and 4. Only for small loop delays in conjunction with a small phase margin is there a significant difference between the characteristics of both
PLLs. For phase margins of practical interest, we can use the simple analytical results of Appendix II as a close approximation for those of the discrete-time PLL.

Since PM depends on the product of \((M + 0.5)\) and \(\omega_n T\), the curves of Fig. 5 (and similarly for Fig. 6) are identical except for a \(y\)-axis scaling factor. Accordingly, for a given phase margin, their global maximum occurs for the same damping factor \(\zeta\), irrespective of \(M\). Similarly, the minimum damping factor \(\zeta_{\text{min}}\) that is needed to achieve a prescribed phase margin PM is independent of \(M\) and is the solution of the equation \(\text{PM} = \arctan[2\pi F(\zeta)]\). An equivalent (though slightly more complex) equation was derived earlier for loops without delay in [9, eq. 6.31]. Fig. 7 depicts \(\zeta_{\text{min}}\) for phase margins between \(0^\circ\) and \(90^\circ\). For small phase margins (say below \(45^\circ\)), \(\zeta_{\text{min}}\) increases linearly with PM. At higher phase margins, it increases ever more rapidly, and very large damping factors are required for phase margins close to \(90^\circ\).

Fig. 8, computed from (2), depicts the maximum normalized loop delay \(\lambda\omega_n = (M + 0.5)\omega_n T\) that is permissible for a given phase margin PM.

For a practical damping factor in the order of unity or somewhat higher, the normalized loop delay should apparently be of the order of 0.1 to 0.2 to achieve practical phase margins on the order of \(45^\circ\) to \(60^\circ\). The points at which the curves of Fig. 8 cross the \(y\) axis define the highest phase margins that are achievable with a given damping factor and are consistent with the minimum damping factor that is needed for a prescribed phase margin as in Fig. 7.

V. FINAL REMARKS

For loop conditions of practical interest, we have found that the discrete-time and continuous-time loops behave essentially identically in the presence of a loop delay. This is true both for first-order and second-order loops. The above results permit us to formulate the following simple rules of thumb.

1) In first-order loops, loop delay \(M\) should be less than half of the normalized loop time constant \(\tau/T = 1/K_t\) in order for the loop to have an adequate phase margin. Equivalently, loop dynamics should be dimensioned for \(\tau/T\) to be at least twice as large as \(M\).

2) In high-gain second-order loops, the damping factor \(\zeta\) is preferably selected somewhat larger than unity, irrespective of \(M\). Here, loop delay \(M\) should be at least 5 to 10 times smaller than the inverse of the normalized natural frequency \(\omega_n T\) in order for the loop to have an adequate phase margin. Equivalently, \(\omega_n T\) should be chosen to be at least 5 to 10 times smaller than \(1/M\).

These rules are likely to be helpful in the design of the concerned loops.

APPENDIX I

PHASE MARGIN OF HIGH-GAIN SECOND-ORDER DISCRETE-TIME PLL

The loop of Fig. 2 has transfer function

\[
G(z) = \frac{\Delta(z)}{\psi_i(z)} = \frac{1}{1 + H(z)}
\]

with \(H(z) = e^{-zM}[K_p^t + K_f^t I(z)] I(z)\). In frequency-domain notation, we may write

\[
I(e^{j\omega T}) = \frac{j e^{-j\pi\Omega}}{2 \sin \pi\Omega} = \frac{1}{2} [1 + j \cot \pi\Omega]
\]

and

\[
K_p^t + K_f^t I(e^{j\omega T}) = K_p^t - \frac{K_f^t}{2} [1 + j \cot \pi\Omega] = A(e^{j\pi\Omega}) e^{j\alpha (e^{j\omega T})}
\]

where

\[
A^2(e^{j\omega T}) = \left( K_p^t - \frac{K_f^t}{2} \right)^2 + \left( \frac{K_f^t}{2} \cot \pi\Omega \right)^2 = K_p^t \left[ K_p^t - K_f^t \right] + \frac{K_f^t}{4 \sin^2 \pi\Omega}
\]

and

\[
\alpha(e^{j\omega T}) = \arctan \left[ \frac{K_f^t \cot \pi\Omega}{K_f^t - 2K_p^t} \right].
\]
Clearly, \[ H(e^{j2\pi\Omega}) = A(e^{j2\pi\Omega})/(2\sin \pi\Omega) \] and therefore \[ H(e^{j2\pi\Omega}) = 1 \] if and only if \[ A(e^{j2\pi\Omega}) = 4\sin^2 \pi\Omega \], i.e., if
\[
(K_p^t)^2 - K_p^t K_f^t + \frac{K_f^t}{4\sin^2 \pi\Omega} = 4\sin^2 \pi\Omega
\]
or, equivalently, if
\[
x^2 - K_p^t [K_p^t - K_f^t] x - K_f^t = 0
\]
where \( x = 4\sin^2 \pi\Omega \). The solution of this equation is
\[
x_{1,2} = \frac{K_f^t}{2} [K_p^t - K_f^t] \pm \frac{1}{2}\sqrt{q}
\]
where \( q = (K_p^t[K_p^t - K_f^t])^2 + 4(K_f^t)^2 \). This may alternatively be denoted
\[
x_{1,2} = \beta \pm \sqrt{\beta^2 + (K_f^t)^2}
\]
(3)
where
\[
\beta = \frac{K_f^t}{2} [K_p^t - K_f^t].
\]
For practical values of \( K_p^t \) and \( K_f^t \), \( \beta \) will always be positive, and, since the desired value of \( x \) is also positive, the solution will be \( x_2 \), i.e., the root with the + sign in (3).

It should be noted that \( x_2 \) is four times the square of a sine, and for this reason \( x \) can fundamentally not become larger than 4. In cases where \( x_2 \) exceeds 4, there exists no frequency \( \Omega_0 \) for which \( |H| = 1 \) for the given values of \( K_p^t \), \( K_f^t \), and \( M \).

The phase margin is determined by the phase \( H \) at the frequency \( \Omega_0 \) that was just identified. The phase \( \phi(e^{j2\pi\Omega}) \) of \( H(e^{j2\pi\Omega}) \) depends on \( \alpha(e^{j2\pi\Omega}) \) according to
\[
\phi(e^{j2\pi\Omega}) = -2\pi M \Omega + \alpha(e^{j2\pi\Omega}) - \pi - \frac{\pi}{2}.
\]
(4)
It follows that
\[
\text{PM} = \frac{\pi}{2} + \arctan \left[ \frac{K_f^t \cot \pi\Omega}{K_f^t - 2K_p^t} \right] - (2M + 1)\pi\Omega_0
\]
(5)
\[
= \arctan \left[ \frac{2K_f^t - K_p^t}{K_f^t \cot \pi\Omega_0} \right] - (2M + 1)\pi\Omega_0.
\]

APPENDIX II

PHASE MARGIN OF HIGH-GAIN SECOND-ORDER CONTINUOUS-TIME PLL

The model of this PLL is the one of Fig. 2 but with all discrete-time operations replaced by their continuous-time counterparts. Specifically, discrete-time integrators are replaced by continuous-time integrators (with transfer function \( 1/(j\omega T) \)) where \( \omega \) is the angular frequency, and the discrete loop delay of \( M \) sampling intervals is replaced by a continuous delay \( \lambda = (M + 0.5)T \). The corresponding open-loop transfer function \( H(j\omega) \) is
\[
H(j\omega) = e^{-j\omega(M+0.5)T} \left[ \frac{K_f^t + K_p^t}{j\omega T} \right] \frac{1}{j\omega T}.
\]

We first identify the angular frequency \( \omega_0 \) at which \( |H| = 1 \). Clearly, \( |H(j\omega)|^2 = [(K_f^t)^2 + (K_p^t/\omega T)^2]/(\omega T)^2 \) so that \( |H| = 1 \) if and only if
\[
(\omega T)^4 - (\omega/\omega_0)^2(K_f^t)^2 - (K_p^t)^2 = 0.
\]
We recall that \( K_p^t = 2\omega_0 T \) and \( K_f^t = (\omega_0 T)^2 \). Condition (6) may be recast in terms of \( \omega_0 \) and \( \zeta \) as
\[
\left( \frac{\omega}{\omega_0} \right)^4 - 4\zeta^2 \left( \frac{\omega}{\omega_0} \right)^2 - 1 = 0.
\]
(7)

It can be observed that only the ratio of \( \omega \) and \( \omega_0 \) comes into play. The absolute value of \( \omega_0 \) does not matter. Equation (7) has only one positive root, which is given by
\[
\left( \frac{\omega_0}{\omega} \right)^2 = 2\zeta^2 + \sqrt{1 + 4\zeta^4}.
\]

Correspondingly, \( \omega_0 = \omega_0 F(\zeta), \) where \( F(\zeta) = \sqrt{2 \zeta^2 + \sqrt{1 + 4 \zeta^4}} \).

Having identified \( \omega_0 \), we next turn our attention to the phase characteristics \( \phi(j\omega) \) of \( H(j\omega) \). Clearly
\[
\phi(j\omega) = -\omega(M + 0.5)T - \arctan \left( \frac{K_f^t}{\omega K_p^t} \right) - \frac{\pi}{2}
\]
\[
= -\omega(M + 0.5)T - \arctan \left( \frac{1}{\omega_0} \right) - \frac{\pi}{2}.
\]
(8)
The phase margin may be expressed in terms of \( \phi \) and \( \omega_0 \) as
\[
\text{PM} = \pi + \phi(\omega_0) = \arctan \left[ 2\zeta F(\zeta) \right] - F(\zeta)(M + 0.5)\omega_0 T.
\]

REFERENCES