

8.1. Az aritmetikai zaj

Ebben a fejezetben a normál működésű, fixpontos aritmetikai egységekben fellépő effektussal fogunk foglalkozni. A véges szóhosszúság következtében a számítási pontosság véges, ami úgy jelentkezik, hogy a valóságos kimenet el fog térni az ideális (a végtelen szóhosszúsághoz tartozó) kimeneti értéktől. Ez az eltérés felfogható úgy is, mint az aritmetikai egység zaja. Ez a zaj különösen jelentős lehet a rekurzív rendszerekben (amikor a kerekítésekkel elkövetett hiba visszacsatolódik)

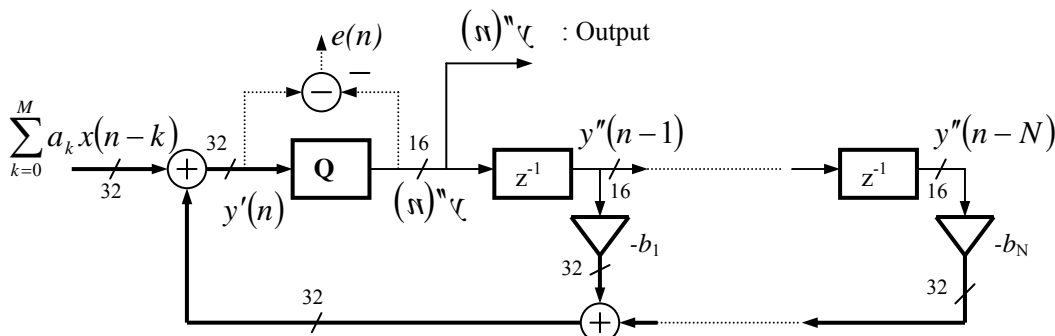
Vizsgálatunkat kezdjük a D0 struktúrájú ARMA rendszerrel, melyet lineáris esetben (végtelen szó-hosszúság esetén) az alábbi differencia egyenlet ír le:

$$y(n) = \sum_{k=0}^M a_k x(n-k) - \sum_{k=1}^N b_k y(n-k) \quad (8.3.)$$

A normál működés azt jelenti, hogy túlvezérlés nem lép fel, ugyanakkor a bemenetre érkező véletlen változónak tekintett $x(n)$ jel szórása a q kvantálási lépcsőnél jóval nagyobb.

A számításokban $x(n)$ pontosságával itt nem kell foglalkoznunk, mert a bemenő jel kvantálásánál (lásd a 7. fejezet) azt már figyelembe vettük. (Az $x(n)$ a tényleges műsorjel és az annak kvantálásakor keletkező zaj összege.)

A szűrőben a szóhosszúság alakulását a 8.2. ábra szemlélteti:



8.2. ábra Az aritmetikai zaj eredete fixpontos aritmetikában

Jelölje $y''(n-k)$ a korábbi ütemek 16 bites eltárolt mintáit és $y'(n)$ az n -ik ütemben az akkumulátor 32 bites tartalmát. Ezzel:

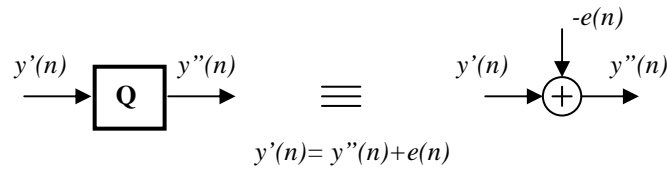
$$y'(n) = y''(n) + e(n) = \sum_{k=0}^M a_k x(n-k) - \sum_{k=1}^N b_k y''(n-k) \quad (8.4.)$$

Kerekítés után a 16 bites $y''(n)$ -et mentjük el. A kerekítés (a kvantálás) hibája:

$$-\frac{q}{2} \leq e(n) < \frac{q}{2} \quad (8.5.)$$

ahol q az LSB helyiértéke (16 TC Q15 kód esetén $q = 2^{-15}$).

A (8.4.) egyenletben az $y'(n)$ felbontása $y''(n)$ -re és $e(n)$ -re, lehetővé teszi a modell linearizálását. Ez a felbontás azt jelenti, hogy a nemlineáris kvantálót helyettesítjük egy additív típusú zajforrással (lásd 8.3. ábra).



8.3. ábra A kvantáló additív zajmodellje

A modellben feltételezzük, hogy az $e(n)$ kvatálási hiba egy véletlen változó jel, amelyik korrelálatlan, független a bemenő jeltől és egyenletes eloszlású a $[-q/2, +q/2]$ tartományban. A 7. fejezet alapján mondhatjuk, hogy ez a modell a feldolgozandó jelek széles osztályára nagyon jó közelítést jelent.

A zaj a keletkezés helyén tehát fehér zaj, melynek spektrális teljesítmény sűrűség függvénye konstans:

$$S_{ee}(\omega) = \frac{q^2}{12} \quad (8.6.)$$

Az $f(n)$ kimeneti aritmetikai zajt az ideális és a valóságos kimenet különbségként értelmezzük.

$$f(n) = y(n) - y''(n) \quad (8.7.)$$

A (8.3.) és a (8.4.) egyenletek különbségét képezve:

$$y(n) - y'(n) = f(n) - e(n) = -\sum_{k=1}^N b_k y(n-k) - \sum_{k=1}^N b_k y''(n-k) = -\sum_{k=1}^N b_k f(n-k) \quad (8.8.)$$

amiből:

$$e(n) = \sum_{k=0}^N b_k f(n-k) \quad \text{ahol: } b_0 = 1 \quad (8.9.)$$

Képezzük a (8.9.) sorozat "Z" transzformáltját:

$$E(z) = F(z) [1 + b_1 z^{-1} + \dots + b_N z^{-N}] = F(z) B(z) \quad (8.10.)$$

ahol:

$$B(z) = 1 + b_1 z^{-1} + \dots + b_N z^{-N} \quad (8.11.)$$

D0 struktúra esetén a zaj keletkezési helyétől a kimenetig a transzfer függvény:

$$G(z) = \frac{F(z)}{E(z)} = \frac{1}{B(z)} \quad (8.12.)$$

A kimeneten az aritmetikai zaj spektrális teljesítmény sűrűség függvénye:

$$S_{ff}(z) = G(z) G(z^{-1}) S_{ee}(z) = \frac{q^2}{12} G(z) G(z^{-1}) \quad (8.13.)$$

A frekvencia tartományban:

$$S_{ff}(\omega) = |G(\omega)|^2 S_{ee}(\omega) = \frac{q^2}{12} |G(\omega)|^2 \quad (8.14.)$$

ahol:

$$G(\omega) = G(z = e^{j\omega T}) \quad (8.15.)$$

A kimeneti aritmetikai zaj átlagteljesítményét az autokorrelációs függvényének zérus helyen felvett értékeként határozzuk meg:

$$\sigma_f^2 = E\{f^2(n)\} = R_{ff}(0) \quad (8.16.)$$

A Wiener-Hincsin tétel értelmében az autokorrelációs függvény a spektrális teljesítmény függvény inverz "Z" transzformáltjaként számítható:

$$R_{ff}(m) = \frac{1}{j2\pi} \oint_C S_{ff}(z) z^{m-1} dz \quad (8.17.)$$

ahol a C görbe esetünkben az egységkör.

Speciálisan:

$$\sigma_f^2 = R_{ff}(0) = \frac{1}{j2\pi} \oint_C S_{ff}(z) z^{-1} dz = \sum_k \text{Res}[S_{ff}(z) z^{-1}, p_k] \quad (8.18.)$$

A C görbére vett integrál a Cauchy féle residuum tétellel (lásd 8.18.) számolható, ahol p_k az integrandus függvénynek a C görbén (az egységkörön) belül lévő pólusait jelöli. Emlékeztetőül, egyszeres pólusokra:

$$\text{Res}[X(z), p] = \lim_{z \rightarrow p} (z - p) X(z) \quad (8.19.)$$

A D0 struktúra esetén:

$$S_{ff}(z) z^{-1} = \frac{q^2}{12} \frac{1}{B(z)} \frac{1}{B(z^{-1})} z^{-1} \quad (8.20.)$$

Másodfokú alaptagra:

$$B(z) = 1 + b_1 z^{-1} + b_2 z^{-2} = (1 - p_1 z^{-1})(1 - p_2 z^{-1}) = 1 - (p_1 + p_2) z^{-1} + p_1 p_2 z^{-2} \quad (8.21.)$$

ahol a gyökök és az együtthatók közötti jól ismert összefüggés:

$$b_1 = -(p_1 + p_2) \quad b_2 = p_1 p_2 \quad (8.22.)$$

Másodfokú esetben a kimeneti zaj spektrális teljesítmény sűrűség függvénye:

$$S_{ff}(z) z^{-1} = \frac{q^2}{12} \frac{z}{(z - p_1)(z - p_2)} \frac{1}{(1 - p_1 z)(1 - p_2 z)} \quad (8.23.)$$

A kimeneti zaj átlagteljesítményének meghatározása végett először számítsuk ki a (8.16.)-ban szereplő residuumokat. A (8.23.) szerinti kifejezésnek az egységkörön belül csak két pólusa van (p_1, p_2).

$$\operatorname{Res}\left[S_{ff}(z)z^{-1}, p_1\right] = \frac{q^2}{12} \frac{p_1}{(p_1 - p_2)} \frac{1}{(1 - p_1^2)(1 - p_1 p_2)} \quad (8.24.)$$

$$\operatorname{Res}\left[S_{ff}(z)z^{-1}, p_2\right] = \frac{q^2}{12} \frac{p_2}{(p_2 - p_1)} \frac{1}{(1 - p_2^2)(1 - p_1 p_2)} \quad (8.25.)$$

A két residuumot behelyettesítve (8.18)-ba kapjuk:

$$\sigma_f^2 = \frac{q^2}{12} \frac{1 + p_1 p_2}{1 - p_1 p_2} \frac{1}{(1 - p_1^2)(1 - p_2^2)} = \frac{q^2}{12} \frac{1 + p_1 p_2}{1 - p_1 p_2} \frac{1}{(1 + p_1 p_2)^2 - (p_1 + p_2)^2} \quad (8.26.)$$

Felhasználva (8.22)-t, a kimeneti zaj teljesítménye:

$$\sigma_f^2 = \frac{q^2}{12} \frac{1 + b_2}{1 - b_2} \frac{1}{(1 + b_2)^2 - b_1^2} \quad (8.27.)$$

Elsőfokú alaptagra ($b_2=0$):

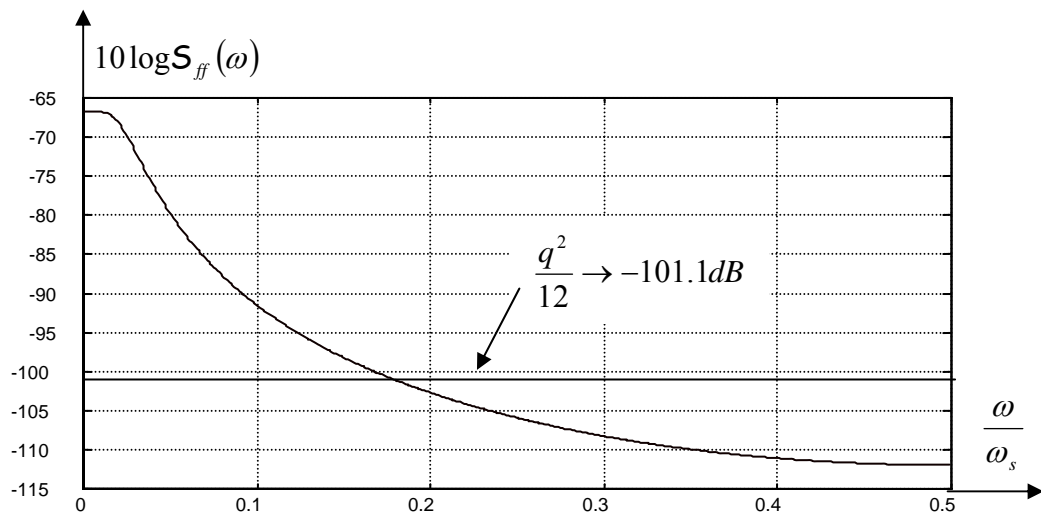
$$\sigma_f^2 = \frac{q^2}{12} \frac{1}{1 - b_1^2} \quad (8.28.)$$

A kvantálási hiba visszacsatolása jelentősen megnövelheti a kimeneti zaj értékét.

Példaként nézzük azt az esetet, mikor $p_{1,2} = 0.9 \pm j0.1$. Ekkor $b_1 = -1.8$ és $b_2 = 0.82$, amivel:

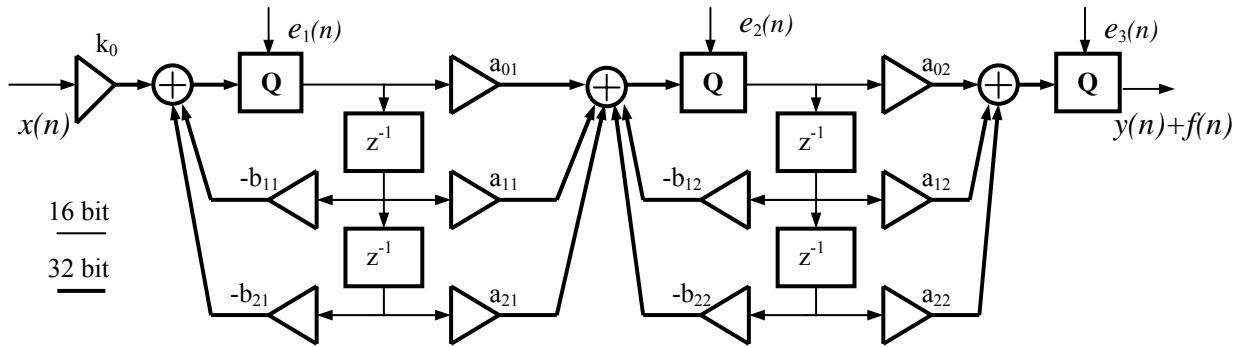
$$\sigma_f^2 = \frac{q^2}{12} * \frac{1.82}{0.18} * \frac{1}{1.82^2 - 1.8^2} = \frac{q^2}{12} 139.65$$

Láthatóan a felszorzódás jelentős, 139.65 –szörös!



8.4. ábra A $p_{1,2} = 0.9 \pm j0.1$ pólusokhoz tartozó kimeneti aritmetikai zaj spektrális sűrűségfüggvénye

Az additív zajmodell alkalmazására tekintsünk még egy példát. A szűrő legyen egy negyedfokú, elliptikus aluláteresztő szűrő, melyet másodfokú D1 struktúrájú másodfokú alaptagok kaszkád kapcsolásával realizálunk (lásd 8.5. ábra).



8.5. ábra Negyedfokú, D1 struktúrájú szűrő zajforrásai

A szűrő transzfer függvénye:

$$H(z) = k_0 H_1(z) H_2(z) \quad (8.29.)$$

ahol:

$$H_1(z) = \frac{a_{01} + a_{11}z^{-1} + a_{21}z^{-2}}{1 + b_{11}z^{-1} + b_{21}z^{-2}} \quad H_2(z) = \frac{a_{02} + a_{12}z^{-1} + a_{22}z^{-2}}{1 + b_{12}z^{-1} + b_{22}z^{-2}} \quad (8.30.)$$

A lehetséges szorzás-akkumulálás funkciókat összevonva, a kvantálások számát minimalizálhatjuk. Esetünkben ez a szám, a zajforrások száma: 3. A zaj forrásaitól a kimenetig érvényes transzfer függvények:

$$G_1(z) = H_1(z) H_2(z), \quad G_2(z) = H_2(z) \quad \text{és} \quad G_3(z) = 1 \quad (8.31.)$$

A zajforrásokat egymástól függetlennek tételezve fel, a kimeneti teljesítmény sűrűség függvények összeadódnak:

$$S_{ff}(z) = S_{ff1}(z) + S_{ff2}(z) + S_{ff3}(z) = [G_1(z)G_1(z^{-1}) + G_2(z)G_2(z^{-1}) + 1] \frac{q^2}{12} \quad (8.32.)$$

Az ábrán látható k_0 szorzóra a szűrő túlvezérlésének elkerülése érdekében van szükség. Megfelelő megválasztásával az első kaszkád fokozat tárolóinak túlvezérlése kerülhető el.

Az első zajforráshoz tartozó transzfer függvény átvitele:

$$G_1(z) = H_1(z) H_2(z) = \frac{H(z)}{k_0} \quad (8.33.)$$

Az áteresztő tartományban (ahol $H(\omega) \approx 1$) és ha $k_0 \ll 1$, akkor $G_1(\omega) \gg 1$, ami szavakban azt jelenti, hogy a kimeneti zaj a nagy erősítés miatt jelentős lesz.

Ezt elerülendő k_0 -t a lehető legnagyobbra kell választani. A szűrő skálázását (a túlvezérlés elkerülését) tehát úgy kell elvégezni, hogy a realizált pólusok és a zérusok sorrendjének megválasztásával a szűrő erősítését a lehető legmagasabb szinten tartjuk, így a hátrébb lévő fokozatoknak nem kell feleslegesen nagyot erősíteni. (Az eredő átvitel előírt!)

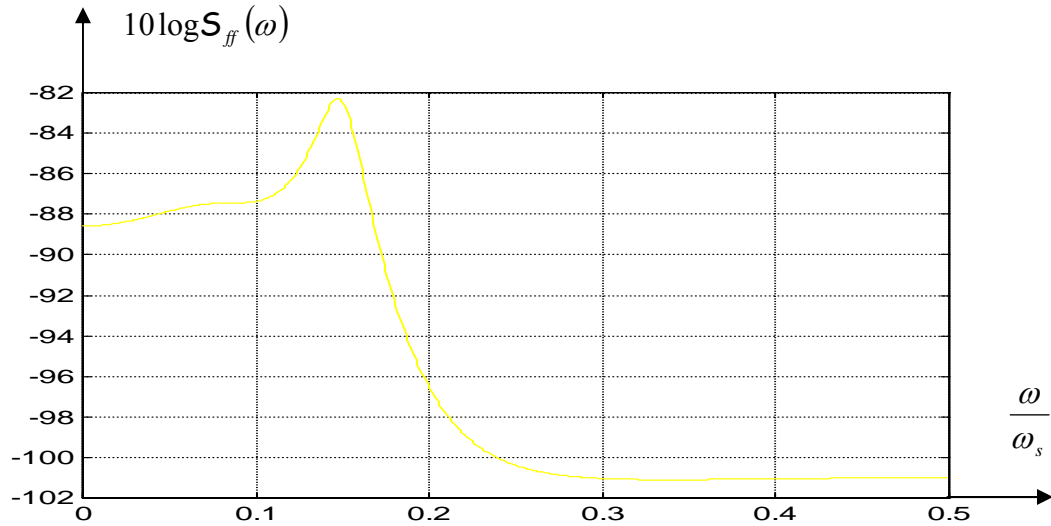
Példánkban a fentebb leírt szempontok szerint skálázva a szűrőt az együtthatók az alábbiaknak adódtak:

$$k_{\sigma} = 0.2402$$

$$a_{01} = 0.3378 \quad a_{11} = -0.1323 \quad a_{21} = 0.3378 \quad b_{11} = -1.2368 \quad b_{21} = 0.4933$$

$$a_{02} = 0.4352 \quad a_{12} = 0.4584 \quad a_{22} = 0.4352 \quad b_{12} = -1.0842 \quad b_{22} = 0.8430$$

A teljes kimeneti zajspektrum a 8.6. ábrán látható.



8.6. ábra A kimeneti aritmetikai zaj spektrális sűrűségfüggvénye

Ha a spektrális sűrűségfüggvény numerikusan ismert (MATLAB), akkor numerikus integrálással számíthatjuk ki legegyszerűbben a kimeneti zaj teljesítményét:

$$\sigma_f^2 = \frac{1}{\omega_s} \int_{-\frac{1}{2}\omega_s}^{\frac{1}{2}\omega_s} S_{ff}(\omega) d\omega = \frac{2}{\omega_s} \int_0^{\frac{1}{2}\omega_s} S_{ff}(\omega) d\omega \approx \frac{2}{\omega_s} \sum_{i=1}^N S_{ff}(\omega_i) \frac{\omega_s}{2N} = \frac{1}{N} \sum_{i=1}^N S_{ff}(\omega_i) \quad (8.34.)$$

$$\text{ahol:} \quad \omega_i = i \frac{\omega_s}{2N} \quad \text{és} \quad \Delta\omega = \frac{\omega_s}{2N} \quad (8.35.)$$

A fenti példára kiszámítva (8.34.)-et

$$\sigma_f^2 = 11.08 * \frac{q^2}{12} \quad (8.36.)$$

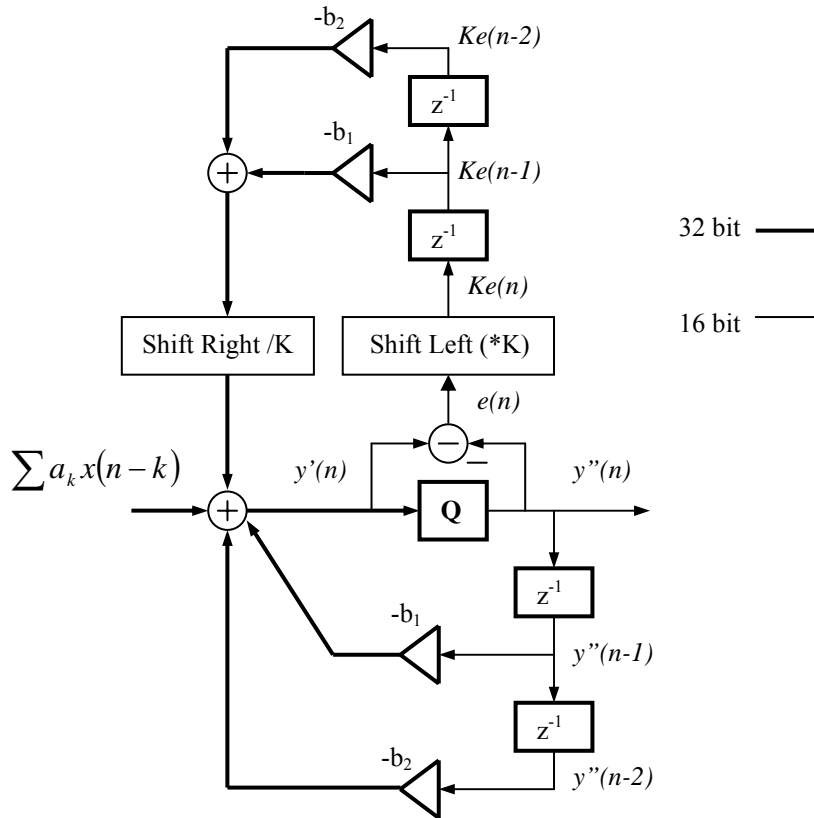
adódik.

Bizonyos alkalmazásokban zavaró lehet az aritmetikai zaj magas szintje. Felmerül a kérdés, hogyan lehetne csökkenteni a zaj értékét?

Egy megoldás lehet a szóhosszúság növelése (praktikusan DSP-k esetében a szóhosszúság kétszerezése). Ez általában kényelmetlen (hosszabb a kód). Azonban nem feltétlen szükséges a teljes szűrőt duplapontosan számolni, néha elég csak a zaj szempontjából kritikus fokozatot. Ez a megoldás ajánlott a zérus bemenetű határosszcilláció megszüntetésére is (lásd később).

A duplapontos számítás esetünkben a hiba visszacsatolásának felel meg. A kvantálási hibát is elmentjük, és hasonlóan processzáljuk mint a “hasznos” jelet.

A kerekítési hiba minta elmentése az akkumulátor alsó helyiértékeiről történik. Az a jel balra történő shiftelésének felel meg. A hibajelek feldolgozása után, annak eredményét kvantálás előtt a megfelelő helyiértéken kell az akkumulátorhoz hozzáadni, azaz jobbra kell eltolni. Ekkor az akkumulátorból lecsordulnak a hibajelből származó eredmény alsó helyiértékei, de ez a hiba már másodrendűen kicsinek számít és ezért ezt elhanyagoljuk.



8.7. ábra A hiba visszacsatolása az aritmetikai zaj csökkentése céljából

A számítás algoritmus:

$$y'(n) = y''(n) + e(n) = \sum_{k=0}^M a_k x(n-k) - \sum_{k=1}^N b_k y''(n-k) - \frac{1}{K} \sum_{k=1}^N b_k Ke(n-k) \quad (8.37.)$$

A (8.3.) és a (8.37.) egyenletek különbségét képezve:

$$y(n) - y'(n) = f(n) - e(n) = -\sum_{k=1}^N b_k f(n-k) + \sum_{k=1}^N b_k e(n-k) \quad (8.38.)$$

amit átrendezve:

$$\sum_{k=0}^N b_k f(n-k) = \sum_{k=0}^N b_k e(n-k) \quad (8.39.)$$

A (8.39.) egyenletet 'Z' transzformálva az:

$$F(z)B(z) = E(z)B(z) \quad (8.40.)$$

összefüggést kapjuk, amiből a kimeneti transzfer függvény:

$$G(z) = \frac{F(z)}{E(z)} = 1 \quad (8.41.)$$

értékűre adódik. Szavakban ez azt jelenti, hogy a kvantálási hiba nem sokszorozódik fel (nem csatolódik vissza a rekurzív rendszerbe). A kapott eredmény megegyezik várakozásunkkal.

Összefoglalva eredményeinket:

Fixpontos aritmetikai egységekben a szorzások megnövelik a szóhosszúságot és az eredménynek a memóriába történő elmentése elkerülhetetlenné teszi a szóhosszúság visszaállítását. A szóhosszúság redukciója (a kvantálás) zajossá teszi a számítást.

Ha a jel szórása jóval nagyobb a kvantálási lépcsőnél (q), akkor a kvantálási hiba véletlen változó jelnek tekinthető, amelyik függetlennek tekinthető a kvantálandó jeltől, spektrális sűrűségfüggvénye "fehér" és egyenletes eloszlású.

Ha egy rendszerben több kvantálást kell végezni, akkor ez azt jelenti, hogy több zajforrásunk is van. Ezért arra kell törekedni a számítási algoritmus kialakításánál, hogy a szóhosszúság redukcióját az eredménynek a memóriába történő visszairáskor végezzük csak el, részeredményt lehetőleg ne kerekítsünk.

Több zajforrás esetén az egyedi forrásokat egymástól függetlennek tekintjük, ezért a kimeneten ezek spektrális sűrűségfüggvényei összeadhatók. Az egyedi sűrűségfüggvényeket az egyedi zajforrásoktól a kimenetig terjedő transzfer függvények segítségével határozzuk meg. A kimeneti teljesítmény sűrűség függvények a transzfer függvények abszolút érték négyzete szerint "színezik" el a forrás "fehér" zaját.

A rendszerek aritmetikai zajának analízise szorosan összefügg az erősítés szintek beállításával (a skálázással). Megállapítottuk, hogy az aritmetikai zaj akkor tartható relatíve alacsony szinten, ha a zajforrás után nincs nagy erősítésű fokozat. Kaszkád realizálás esetén ezért rendszertechnikai megfontolás tárgyává kell tenni a fokozatok sorrendjének meghatározását.

Bizonyos speciális bemenő jelekre, melyekre nem teljesül a bemenő jel szórására tett fenti feltevésünk, akkor természetesen az aritmetikai zajmodell nem megfelelő. Ilyen jel lehet a konstans bemenet, vagy az az eset mikor bizonyos idejű működés után a bemenő jelet kikapcsoljuk (zérus bemenet) és a kikapcsolási tranziensre vagyunk kíváncsiak. A kvantálás okozta nemlinearitás stabilitási problémát jelenthet. Ezekkel a kérdésekkel külön fejezetben foglalkozunk.