

Estimatie en detectie op basis van signalen met onbekende parameters

Strategieën

In de context van digitale communicatie bevat het uitgezonden signaal, naast de te detecteren digitale informatie, verschillende parameters (zoals amplitude, fase, frequentie en tijdsvertraging) waarvan de waarde niet is gekend. We zullen ons in hoofdzaak concentreren op het geval van lineaire digitale communicatie.

Laten we de te detecteren datasymbolen en de onbekende parameters noteren als respectievelijk \mathbf{a} en \mathbf{x} , en de gezamenlijke kansfunctie van \mathbf{a} en \mathbf{x} als $p(\mathbf{r}|\mathbf{a},\mathbf{x})$. Waar nodig onderstellen we dat \mathbf{a} (discreet) en \mathbf{x} (continu) statistisch onafhankelijk zijn met a priori distributies $\Pr[\mathbf{a}=\mathbf{a}]$ en $p(\mathbf{x})$. We kunnen verschillende strategieën volgen om de datasymbolen te detecteren. Een aantal worden hierna besproken.

Strategie 1 : ML detectie

De eerste strategie is het minimaliseren van de foutprobabiliteit $\Pr[\hat{\mathbf{a}} \neq \mathbf{a}]$, wat zich herleidt tot de ML decisieregel wanneer de mogelijke symboolsequenties dezelfde a priori waarschijnlijkheid hebben. De decisieregel is dan $\hat{\mathbf{a}}_{ML} = \arg \max_{\mathbf{a}} p(\mathbf{r}|\mathbf{a})$, waarbij $p(\mathbf{r}|\mathbf{a})$ de kansfunctie van de symboolsequentie voorstelt. Deze kansfunctie is gegeven door

$$p(\mathbf{r}|\mathbf{a}) = \int p(\mathbf{r}|\mathbf{a},\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

waarbij $p(\mathbf{x})$ de a priori distributie van de parameter \mathbf{x} voorstelt. Deze strategie heeft het nadeel dat de te maximaliseren kansfunctie $p(\mathbf{r}|\mathbf{a})$ in het algemeen moeilijk te berekenen is, en dat de rekencomplexiteit m.b.t. de maximalisatie van $p(\mathbf{r}|\mathbf{a})$ exponentieel toeneemt met de lengte van de datasequentie (wanneer de lengte van de datasequentie gelijk is aan L en elk datasymbool M waarden kan aannemen, moeten M^L mogelijke datasequenties onderzocht worden). Het toepassen van deze strategie kan enkel indien er benaderingen worden toegepast, die echter hun weerslag hebben op de betrouwbaarheid van de detectie.

Strategie 2 : gezamenlijke ML estimatie en detectie

Een tweede strategie is de gezamenlijke estimatie en detectie van (\mathbf{a},\mathbf{x}) . Het ML criterium leidt tot

$$(\hat{\mathbf{a}}_{ML}, \hat{\mathbf{x}}_{ML}) = \arg \max_{\mathbf{a},\mathbf{x}} p(\mathbf{r}|\mathbf{a},\mathbf{x})$$

Deze maximalisatie kan in principe uitgevoerd op de volgende wijze :

- Bepaal voor alle mogelijke testwaarden $\tilde{\mathbf{x}}$ van de parameter \mathbf{x} de bijhorende decisie $\hat{\mathbf{a}}(\tilde{\mathbf{x}})$

$$\hat{\mathbf{a}}(\tilde{\mathbf{x}}) = \arg \max_{\mathbf{a}} p(\mathbf{r}|\mathbf{a},\tilde{\mathbf{x}})$$

Dit correspondeert met de 'klassieke' detectie van de symbolen \mathbf{a} , waarbij de testwaarde $\tilde{\mathbf{x}}$ wordt beschouwd als de correcte waarde van de parameter \mathbf{x}
- Bepaal $\hat{\mathbf{x}}_{ML}$ als $\hat{\mathbf{x}}_{ML} = \arg \max_{\tilde{\mathbf{x}}} p(\mathbf{r}|\hat{\mathbf{a}}(\tilde{\mathbf{x}}),\tilde{\mathbf{x}})$. Deze optimalisatie heeft in het algemeen geen analytische oplossing, zodat een zoekprocedure is vereist
- Tenslotte bekomen we $\hat{\mathbf{a}}_{ML}$ als $\hat{\mathbf{a}}_{ML} = \hat{\mathbf{a}}(\hat{\mathbf{x}}_{ML})$

Een nadeel van deze tweede strategie is de rekencomplexiteit : bij elke testwaarde $\tilde{\mathbf{x}}$ van \mathbf{x} moet een nieuwe decisie $\hat{\mathbf{a}}(\tilde{\mathbf{x}})$ gemaakt worden.

Strategie 3 : DDML schatting van \mathbf{x} , met initialisatie door trainingsequentie

De rekencomplexiteit van de tweede strategie kan gereduceerd worden door gebruik te maken van een trainingsequentie. We delen de datasequentie op in blokken van L symbolen, en om de $N-1$ blokken wordt een trainingsequentie van lengte L toegevoegd. We noteren het i -de symboolblok als $\mathbf{a}^{(i)}$, en de corresponderende observatie en parameter als $\mathbf{r}^{(i)}$ en $\mathbf{x}^{(i)}$. De blokindices van de trainingsequenties zijn $0, N, 2N, \dots$. Op basis van de observatie $\mathbf{r}^{(0)}$ en de gekende trainingsequentie $\mathbf{a}^{(0)}$ wordt een schatting $\hat{\mathbf{x}}^{(0)}$ gemaakt (bijvoorbeeld volgens het ML criterium). In de onderstelling dat de waarde van de parameter \mathbf{x} nauwelijks verandert tussen twee opeenvolgende blokken, wordt $\hat{\mathbf{x}}^{(0)}$ gebruikt als zijnde de correcte waarde van $\mathbf{x}^{(1)}$ om op basis van de observatie $\mathbf{r}^{(1)}$ een decisie $\hat{\mathbf{a}}^{(1)}$ te maken (bij voorbeeld volgens het ML criterium) m.b.t. het datablok $\mathbf{a}^{(1)}$. De gedetecteerde sequentie $\hat{\mathbf{a}}^{(1)}$ wordt dan gebruikt als 'trainingsequentie' (alhoewel wegens occasionele decisiefouten $\hat{\mathbf{a}}^{(1)}$ af en toe verschilt van $\mathbf{a}^{(1)}$) om uit de observatie $\mathbf{r}^{(1)}$ een schatting $\hat{\mathbf{x}}^{(1)}$ te maken van $\mathbf{x}^{(1)}$, die dan wordt aangewend in het volgende blok om uit $\mathbf{r}^{(2)}$ een decisie $\hat{\mathbf{a}}^{(2)}$ m.b.t. $\mathbf{a}^{(2)}$ te maken. Het schatten van een parameter op basis van de decisies noemt men een 'decision-directed (DD)' schatting; deze staat in contrast met de 'data-aided (DA)' schatting voortvloeiend uit een a priori gekende trainingsequentie. Een dergelijke strategie is gevoelig aan propagatie van fouten : wanneer de ruis in een blok (met index i) occasioneel sterk is, en heel wat decisiefouten veroorzaakt, zal de corresponderende DD schatting van de parameter $\mathbf{x}^{(i)}$ onnauwkeurig zijn; deze schatting wordt aangewend voor de detectie van de symbolen in het volgende blok (met index $i+1$), waardoor ook $\hat{\mathbf{a}}^{(i+1)}$ veel fouten kan bevatten. Het periodisch invoegen van een trainingsequentie garandeert dat de foutpropagatie beperkt blijft tot aan het eerstvolgende blok met trainingsymbolen.

Strategie 4 : NDAML schatting van \mathbf{x}

Het toevoegen van trainingsymbolen aan de datasequentie leidt tot een toename van de vereiste bandbreedte (omdat er per tijdseenheid meer symbolen moeten verstuurd worden) en tot een 'verkwisting' van zendvermogen (omdat de trainingsequentie geen informatie bevat over de datasymbolen). In een situatie waarbij de zender de data verstuurt in korte bursts (bijvoorbeeld bij TDMA) in plaats van in een continue stroom, kan een trainingsequentie (die voldoende lang is om een nauwkeurige schatting van de parameter \mathbf{x} op te leveren) een ontoelaatbare reductie van de efficiëntie in bandbreedte en vermogen betekenen.

In dit geval maakt men geen gebruik van een trainingsequentie om de parameter \mathbf{x} te schatten. Om de complexiteit van de tweede strategie te vermijden, schat men \mathbf{x} op basis van een criterium dat geen gebruik maakt van (een decisie met betrekking tot) de datasymbolen; dit resulteert in een 'nondata-aided (NDA)' schatting van \mathbf{x} . Een geschikt criterium is de ML schatting van \mathbf{x} , die de kansfunctie $p(\mathbf{r}|\mathbf{x})$ maximaal maakt : $\hat{\mathbf{x}}_{\text{NDAML}} = \arg \max_{\mathbf{x}} p(\mathbf{r} | \mathbf{x})$, met

$p(\mathbf{r} | \mathbf{x}) = \sum_{\boldsymbol{\alpha}} p(\mathbf{r} | \boldsymbol{\alpha}, \mathbf{x}) \Pr[\mathbf{a} = \boldsymbol{\alpha}]$. Deze schatting wordt dan beschouwd als de echte waarde

van \mathbf{x} om een decisie m.b.t. de datasymbolen te bekomen : $\hat{\mathbf{a}} = \arg \max_{\boldsymbol{\alpha}} p(\mathbf{r} | \boldsymbol{\alpha}, \hat{\mathbf{x}}_{\text{NDAML}})$. De

NDAML van \mathbf{x} schatting is echter minder nauwkeurig dan de DAML schatting, zodat de betrouwbaarheid van de detectie lager is dan bij strategie 3.

Differentiële codering en detectie

We beschouwen het geval waarbij het uitgezonden signaal, naast de te detecteren datasequentie \mathbf{a} , een onbekende fase θ bevat : $r(t) = s(t;\mathbf{a})\exp(j\theta) + w(t)$. We bepalen de kansfunctie $p(\mathbf{r}|\mathbf{a})$ uitgaande van de gezamenlijke kansfunctie $p(\mathbf{r}|\mathbf{a},\theta)$. In de onderstelling dat de energie van $s(t;\mathbf{a})$ gelijk is aan E , ongeacht de datasequentie \mathbf{a} , bekommen we

$$p(\mathbf{r}|\mathbf{a},\theta) = \exp\left(\frac{2}{N_0} \operatorname{Re}[\exp(-j\theta)z(\mathbf{a})]\right), \text{ met } z(\mathbf{a}) = \int r(t)s^*(t;\mathbf{a})dt$$

Onderstellen we dat θ uniform verdeeld is over $(-\pi, \pi)$, dan geldt

$$p(\mathbf{r}|\mathbf{a}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\left(\frac{2}{N_0} \operatorname{Re}[\exp(-j\theta)z(\mathbf{a})]\right) d\theta = I_0\left(\frac{2|z(\mathbf{a})|}{N_0}\right)$$

Hierbij is $I_0(\cdot)$ de gewijzigde Besselse functie van orde 0; deze is gedefinieerd als

$$I_0(v) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(v \cos(\theta - \psi)) d\theta. \text{ Merk op dat in deze definitie de waarde van } \psi \text{ geen rol}$$

speelt, omdat het integrandum periodisch is in θ met periode 2π . De functie $I_0(v)$ is even in v , en is monotoon stijgend in $|v|$. Hieruit volgt dat de ML decisie gegeven is door

$\hat{\mathbf{a}}_{\text{ML}} = \arg \max_{\mathbf{a}} |z(\mathbf{a})|$. Deze decisieregels maakt geen gebruik van (een schatting van) de fase : men spreekt van niet-coherente detectie. Ter vergelijking : de ML decisieregels voor coherente detectie is $\hat{\mathbf{a}}_{\text{ML}} = \arg \max_{\mathbf{a}} \operatorname{Re}[\exp(-j\theta)z(\mathbf{a})]$, waarbij θ de gekende (of geschatte) waarde van de fase voorstelt.

In het geval van twee hypothesen H_0 en H_1 stellen we de twee mogelijke uitgezonden signalen voor door $s_0(t)\exp(j\theta)$ en $s_1(t)\exp(j\theta)$. Het is duidelijk dat de keuze $s_0(t) = s_1(t)\exp(j\phi)$ (bijvoorbeeld antipodale signalen bij $\phi=\pi$) leidt tot een slechte performantie, omdat de ontvanger dan geen onderscheid kan maken tussen enerzijds signaal $s_0(t)$ en fase θ , en anderzijds signaal $s_1(t)$ en fase $\theta+\phi$. Men kan aantonen dat de decisiefoutprobabiliteit toeneemt met $|\rho|$, waarbij ρ de complexe correlatie tussen $s_0(t)$ en $s_1(t)$ voorstelt : $\rho = \int s_0^*(t)s_1(t)dt$. De kleinste foutprobabiliteit treedt dus op bij $\rho=0$, (bijvoorbeeld : $s_0(t)$ en $s_1(t)$ overlappen niet in de tijd, of overlappen niet in de frequentie) en is dan gegeven door $\text{BER} = (1/2)\exp(-E/(2N_0))$ (zonder bewijs). Voor $|\rho|\neq 0$ is de uitdrukking voor de foutprobabiliteit bijzonder ingewikkeld. Merk op dat bij coherente detectie de foutprobabiliteit toeneemt met $\operatorname{Re}[\rho]$. Voor $\operatorname{Re}[\rho] = -1$ geldt dan $\text{BER} = Q(\sqrt{2E/N_0})$ (antipodale signalen), terwijl $\text{BER} = Q(\sqrt{E/N_0})$ voor $\operatorname{Re}[\rho] = 0$ (orthogonale signalen).

Laten we het geval van lineaire modulatie beschouwen : $s(t;\mathbf{a}) = \sqrt{E_s} \sum_k a_k p(t-kT)$. We

onderstellen dat $|a_k| = 1$ (M-PSK) en $\int p^*(t)p(t-mT)dt = \delta_m$, zodat de energie van $s(t;\mathbf{a})$ niet afhangt van de datasequentie. De decisieregels voor niet-coherente detectie wordt dan $\hat{\mathbf{a}}_{\text{ML}} = \arg \max_{\mathbf{a}} \left| \sum_k a_k^* z_k \right|$, waarbij $z_k = \int r(t)p^*(t-kT)dt$ wordt bekomen door $r(t)$ aan te

leggen aan het matched filter met impulsantwoord $p^*(-t)$, en de uitgang te bemonsteren op $t-kT$. Omdat de M-PSK constellatie invariant is onder een rotatie van $2\pi/M$, kan deze decisieregels geen onderscheid maken tussen de sequentie \mathbf{a} en de sequentie $\mathbf{a}\exp(j2\pi/M)$: er treedt dubbelzinnigheid (ambiguity) op. Deze dubbelzinnigheid wordt weggenomen door

differentiële codering toe te passen aan de zenderzijde : de te versturen informatie zit niet vervat in het faseverschil tussen twee opeenvolgende symbolen. Wanneer we L M-PSK symbolen (c_1, c_2, \dots, c_L) wensen over te brengen, versturen we $L+1$ M-PSK symbolen (a_0, a_1, \dots, a_L), waarbij a_0 een willekeurig M-PSK symbool is, en $a_k = c_k a_{k-1}$. Het detecteren van de symbolen (c_1, c_2, \dots, c_L) wordt differentiële detectie genoemd. De decisieregel wordt bekomen door in bovenstaande regen voor niet-coherente detectie a_k te vervangen door $c_k c_{k-1} \dots c_2 c_1 a_0$. Deze decisieregel vereist echter een rekencomplexiteit die evenredig is met M^L , zodat voor praktisch gebruik benaderingen zich opdringen.

De eenvoudigste benadering bestaat er in de decisieregel toe te passen op de opeenvolgende paren van verstuurd symbolen (a_0, a_1), (a_1, a_2), ..., (a_{k-1}, a_k), ... De te maximaliseren functie m.b.t. het paar (a_{k-1}, a_k) is gegeven door (we maximaliseren $|z(\mathbf{a})|^2$ i.p.v. $|z(\mathbf{a})|$) :

$$|z_{k-1} a_{k-1}^* + z_k a_k^*|^2 = |z_{k-1}|^2 + 2 \operatorname{Re}[z_k z_{k-1}^* a_k^* a_{k-1}] + |z_k|^2 = |z_{k-1}|^2 + 2 \operatorname{Re}[z_k z_{k-1}^* c_k^*] + |z_k|^2$$

zodat de decisieregel zich herleidt tot $\hat{c}_k = \arg \max_{\alpha} \operatorname{Re}[z_k z_{k-1}^* \alpha^*]$, waarbij α een constellatie-

punt voorstelt. Deze decisieregel betekent dat \hat{c}_k het constellatiepunt is dat het dichtst bij $z_k z_{k-1}^*$ is gelegen. Bij afwezigheid van ruis geldt $z_k = \sqrt{E_s} a_k$, zodat $z_k z_{k-1}^* = E_s c_k$, en $\operatorname{Re}[z_k z_{k-1}^* \alpha^*]$ maximaal wordt wanneer $\alpha = c_k$. In het geval van BPSK geldt $c_k \in \{-1, 1\}$ zodat de twee mogelijke uitgezonden signalen kunnen voorgesteld worden door

$$s_0(t) = \sqrt{E_b} a_{k-1} p(t - kT + T) - \sqrt{E_b} a_k p(t - kT)$$

$$s_1(t) = \sqrt{E_b} a_{k-1} p(t - kT + T) + \sqrt{E_b} a_k p(t - kT)$$

Hieruit volgt dat $\int s_i^*(t) s_j(t) dt = 2E_b \delta_{i-j}$, zodat $\operatorname{BER} = (1/2) \exp(-E_b/N_0)$. In het geval van M-PSK met $M > 2$ is de uitdrukking voor de BER veel ingewikkelder. Fig. 1 vergelijkt de BER voor BPSK en QPSK, zowel voor coherente als voor differentiële detectie. In het geval van coherente detectie leiden BPSK en QPSK tot dezelfde BER. Differentiële detectie bij BPSK leidt tot een kleine degradatie t.o.v. coherente detectie, terwijl differentiële detectie bij QPSK tot een veel grotere degradatie (ongeveer 3 dB) t.o.v. coherente detectie leidt.

Bij transmissie over radiokanalen kan in bepaalde omstandigheden ('vlakke Rayleigh fading') het ontvangen signaal gemodelleerd worden als $r(t) = A s(t; \mathbf{a}) + w(t)$, waarbij $A = A_I + jA_Q$ een complexe amplitude is; A_I en A_Q zijn statistisch onafhankelijke Gaussiaanse toevalsgrootheden met gemiddelde 0 en variantie σ^2 . We bekomen :

$$p(\mathbf{r} | \mathbf{a}) = \int dudv \frac{1}{2\pi\sigma^2} \exp\left(\frac{-1}{2\sigma^2}(u^2 + v^2)\right) \exp\left(\frac{-1}{N_0}((u^2 + v^2)E - 2 \operatorname{Re}[(u - jv)z(\mathbf{a})])\right)$$

$$= \frac{\sigma_0^2}{\sigma^2} \exp\left(\frac{2|z(\mathbf{a})|^2 \sigma_0^2}{N_0^2}\right) \underbrace{\int dudv \frac{1}{2\pi\sigma_0^2} \exp\left(\frac{-1}{2\sigma_0^2} \left(\left(u - \frac{2\sigma_0^2 z_1(\mathbf{a})}{N_0}\right)^2 + \left(v - \frac{2\sigma_0^2 z_Q(\mathbf{a})}{N_0}\right)^2 \right)\right)}_{=1 \text{ (integraal van gezamenlijke Gaussiaanse verdeling van twee toevalsgrootheden)}}$$

met $\frac{1}{\sigma_0^2} = \frac{1}{\sigma^2} + \frac{2E}{N_0}$. Merk op dat ook hier de decisieregel leidt tot het maximaliseren van

$|z(\mathbf{a})|$. In het geval van twee hypothesen waarbij de corresponderende signalen $s_0(t)$ en $s_1(t)$ orthogonaal zijn (complexe correlatie = 0) en dezelfde energie E hebben, is de BER geconditioneerd op de waarde van $|A|^2$ gegeven door $\operatorname{BER}(|A|^2) = (1/2) \exp(-E|A|^2/(2N_0))$. Uitmiddelen over $|A|^2$ (met $\sigma^2 = 1/2$, zodat $E[|A|^2] = 1$) geeft

$$\begin{aligned} \text{BER} &= \int dudv \frac{1}{2\pi\sigma^2} \exp\left(\frac{-1}{2\sigma^2}(u^2 + v^2)\right) \frac{1}{2} \exp\left(\frac{-E}{2N_0}(u^2 + v^2)\right) \\ &= \frac{1}{2} \left(1 + \frac{E\sigma^2}{N_0}\right)^{-1} = \frac{1}{2} \left(1 + \frac{E}{2N_0}\right)^{-1} \cong \frac{N_0}{E} \end{aligned}$$

waarbij bovenstaande benadering opgaat voor grote E/N_0 . De BER voor BPSK met differentiële detectie wordt bekomen door in bovenstaande formule E te vervangen door $2E_b$: voor grote E_b/N_0 geldt dan $\text{BER} \cong N_0 / (2E_b)$. Merk op dat bij gegeven E_b/N_0 de BER bij transmissie over een kanaal met vlakke Rayleigh fading heel wat hoger is dan bij transmissie over een kanaal zonder fading (met $|A| = 1$): bij afwezigheid van fading daalt BER exponentieel met E_b/N_0 , terwijl bij fading BER nagenoeg omgekeerd evenredig is met E_b/N_0 .

Indien A gekend is (of een nauwkeurige schatting van A is beschikbaar), kan coherente detectie worden toegepast. De decisieregel voor M-PSK is dan $\hat{a}_k = \arg \max_{a_k} \text{Re}[a_k^* A^* z_k]$; in het geval van BPSK is de BER, geconditioneerd op $|A|^2$, gegeven door

$$\text{BER}(|A|^2) = Q\left(\sqrt{\frac{2E_b}{N_0}} |A|^2\right). \text{ Uitmiddeling over } |A|^2 \text{ (vlakke Rayleigh fading) geeft dan}$$

$$\begin{aligned} \text{BER} &= \int dudv \frac{1}{2\pi\sigma^2} \exp\left(\frac{-1}{2\sigma^2}(u^2 + v^2)\right) Q\left(\sqrt{\frac{2E_b(u^2 + v^2)}{N_0}}\right) && \begin{aligned} (u &= r \cos \theta \\ v &= r \sin \theta \\ dudv &= r dr d\theta) \end{aligned} \\ &= \int_{-\pi}^{\pi} d\theta \int_0^{\infty} dr \frac{r}{2\pi\sigma^2} \exp\left(\frac{-r^2}{2\sigma^2}\right) Q\left(r \sqrt{\frac{2E_b}{N_0}}\right) \\ &= \int_0^{\infty} dr \frac{r}{\sigma^2} \exp\left(\frac{-r^2}{2\sigma^2}\right) \underbrace{\int_{r\sqrt{\frac{2E_b}{N_0}}}^{\infty} dx \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)}_{Q\left(r \sqrt{\frac{2E_b}{N_0}}\right)} \\ &= \int_0^{\infty} dx \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) \int_0^{x\sqrt{\frac{N_0}{2E_b}}} d\left(\frac{r^2}{2\sigma^2}\right) \exp\left(\frac{-r^2}{2\sigma^2}\right) \\ &= \int_0^{\infty} dx \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) \left(1 - \exp\left(\frac{-x^2 N_0}{4E_b \sigma^2}\right)\right) \\ &= \frac{1}{2} \left(1 - \left(1 + \frac{N_0}{2E_b \sigma^2}\right)^{-1/2}\right) = \frac{1}{2} \left(1 - \left(1 + \frac{N_0}{E_b}\right)^{-1/2}\right) \cong \frac{N_0}{4E_b} \end{aligned}$$

waarbij bovenstaande benadering opgaat voor grote E/N_0 . Bij transmissie van BPSK over een kanaal met vlakke Rayleigh fading is de BER bij coherente detectie dus nagenoeg een factor 2 lager dan bij differentiële detectie.

NDA estimatie

Algoritmes

We beschouwen het geval van lineaire modulatie, waarbij de fase (θ), de frequentie (F) en de timing (τ) onbekende parameters zijn. Voor de eenvoud onderstellen we dat $|a_k| = 1$ (zoals bij M-PSK en 4-QAM). De gezamenlijke kansfunctie van de datasymbolen \mathbf{a} en de onbekende parameters θ, F, τ is gegeven door

$$p(\mathbf{r} | \mathbf{a}, \theta, F, \tau) = \exp\left(\frac{2\sqrt{E_s}}{N_0} \sum_k \operatorname{Re}[a_k^* u_k(\theta, F, \tau)]\right), \text{ met } u_k(\tau) = z_k(\tau) \exp(-j(2\pi FkT + \theta))$$

Hierbij wordt $z_k(\tau)$ bekomen door het ontvangen signaal $r(t)$ aan te leggen aan de ingang van het matched filter met impulsantwoord $p^*(-t)$, en het resulterend signaal te bemonsteren op $t = kT + \tau$. Stilzwijgend hebben we bij de bepaling van de gezamenlijke kansfunctie ondersteld dat $\exp(j2\pi FkT)$ weinig varieert over de duur van de zenderpuls $p(t)$. De kansfunctie van de onbekende parameters wordt bekomen door $p(\mathbf{r} | \mathbf{a}, \theta, F, \tau)$ uit te middelen over \mathbf{a} .

In het geval van BPSK ($a_k \in \{-1, 1\}$) met statistisch onafhankelijke even waarschijnlijke symbolen leidt dit tot

$$p(\mathbf{r} | \theta, F, \tau) = \prod_k E_{a_k} \left[\exp\left(\frac{2\sqrt{E_s}}{N_0} a_k \operatorname{Re}[u_k(\theta, F, \tau)]\right) \right] = \prod_k \cosh\left(\frac{2\sqrt{E_s}}{N_0} \operatorname{Re}[u_k(\theta, F, \tau)]\right), \text{ zodat}$$

$$\ln p(\mathbf{r} | \theta, F, \tau) = \sum_k \ln \cosh\left(\frac{2\sqrt{E_s}}{N_0} \operatorname{Re}[u_k(\theta, F, \tau)]\right). \text{ Het maximaliseren van } \ln p(\mathbf{r} | \theta, F, \tau) \text{ kan}$$

niet analytisch, en vereist kennis van E_s en N_0 . Dit probleem kan omzeild worden door de logaritmische kansfunctie te benaderen door de limiet voor grote N_0 . Rekening houdend met

$$\ln \cosh(v) \cong \ln\left(1 + \frac{v^2}{2}\right) \cong \frac{v^2}{2} \text{ voor } |v| \ll 1 \text{ bekomen we, op irrelevante factoren na,}$$

$$\begin{aligned} \ln p(\mathbf{r} | \theta, F, \tau) &\cong \sum_k (\operatorname{Re}[u_k(\theta, F, \tau)])^2 \\ &= \frac{1}{2} \operatorname{Re} \left[\exp(-j2\theta) \sum_k z_k^2(\tau) \exp(-j4\pi FkT) \right] + \frac{1}{2} \sum_k |z_k(\tau)|^2 \end{aligned}$$

Deze benaderde uitdrukking voor de logaritmische kansfunctie wordt dan gebruikt voor alle N_0 . De maximalisatie van deze benaderde uitdrukking leidt tot schattingen die voldoen aan

$$\begin{aligned} \hat{\theta}_{\text{NDAML}} &= \frac{1}{2} \arg \left(\sum_k z_k^2(\hat{\tau}_{\text{NDAML}}) \exp(-j4\pi \hat{F}_{\text{NDAML}} kT) \right) \\ (\hat{F}_{\text{NDAML}}, \hat{\tau}_{\text{NDAML}}) &= \arg \max_{F, \tau} \left(\left| \sum_k z_k^2(\tau) \exp(-j4\pi FkT) \right|^2 + \sum_k |z_k(\tau)|^2 \right) \end{aligned}$$

Het bepalen van $(\hat{F}_{\text{NDAML}}, \hat{\tau}_{\text{NDAML}})$ vereist een tweedimensionele zoekprocedure, waarna $\hat{\theta}_{\text{NDAML}}$ wordt bekomen door rechtstreekse berekening.

De (tijdrovende !) tweedimensionele zoekprocedure m.b.t. de frequentie en de timing kan vermeden worden door de schatting van de timing te baseren op het maximaliseren van $p(\mathbf{r} | \tau)$,

en vervolgens deze schatting te beschouwen als de correcte waarde van de timing bij het gezamenlijk schatten van (θ, F) . Merk op dat $p(\mathbf{r}|\tau)$ wordt bekomen door $p(\mathbf{r}|\mathbf{a}, \theta, F, \tau)$ uit de middelen over (\mathbf{a}, θ, F) . We zullen deze uitmidding uitvoeren in de onderstelling dat N_0 groot is, zodat we $p(\mathbf{r}|\mathbf{a}, \theta, F, \tau)$ kunnen benaderen door een afgebroken Taylorreeks. Steunend op

$\exp(v) \cong 1 + v + \frac{v^2}{2}$ voor $|v| \ll 1$ bekomen we (voor BPSK)

$$p(\mathbf{r} | \mathbf{a}, \theta, F, \tau) \cong 1 + \left(\frac{2\sqrt{E_s}}{N_0} \sum_k a_k \operatorname{Re}[u_k(\theta, F, \tau)] \right) + \frac{1}{2} \left(\frac{2\sqrt{E_s}}{N_0} \sum_k a_k \operatorname{Re}[u_k(\theta, F, \tau)] \right)^2$$

Rekening houdend met $E[a_k] = 0$ en $E[a_m a_n] = \delta_{m-n}$ leidt de uitmidding over \mathbf{a} tot de eerder bekomen benadering voor $\ln p(\mathbf{r} | \theta, F, \tau)$, die we hadden bekomen door eerst de uitmidding over \mathbf{a} exact te berekenen, en daarna de benadering voor grote N_0 door te voeren. Eerst benaderen en dan uitmiddelen biedt het voordeel dat slechts enkele momenten (i.p.v. de volledige distributie) van de symbolen moeten gekend zijn. Rekening houdend met $E[\exp(-j\theta)] = 0$ leidt de uitmidding van $\ln p(\mathbf{r} | \theta, F, \tau)$ over θ tot (irrelevante factoren

weggelaten) $\ln p(\mathbf{r} | F, \tau) \cong \sum_k |z_k(\tau)|^2$. We merken op dat deze uitdrukking niet meer van F afhangt (de afhankelijkheid van F zit in de hogere-ordetermen van de Taylorreeks van $p(\mathbf{r} | \mathbf{a}, \theta, F, \tau)$, die we hebben verwaarloosd) zodat $\ln p(\mathbf{r} | \tau) \cong \sum_k |z_k(\tau)|^2$. De schattingen van θ , F en τ zijn dan gegeven door

$$\hat{\theta}_{\text{NDAML}} = \frac{1}{2} \arg(L_F(\hat{F}_{\text{NDAML}})) \quad \hat{F}_{\text{NDAML}} = \arg \max_{\hat{F}} |L_F(\hat{F})| \quad \hat{\tau}_{\text{NDAML}} = \arg \max_{\hat{\tau}} L_\tau(\hat{\tau})$$

waarbij $L_F(\hat{F}) = \sum_k z_k^2(\hat{\tau}_{\text{NDAML}}) \exp(-j4\pi\hat{F}kT)$ en $L_\tau(\hat{\tau}) = \sum_k |z_k(\hat{\tau})|^2$. Eerst gebeurt een ééndimensionale zoekprocedure m.b.t. de timing, vervolgens een ééndimensionale zoekprocedure m.b.t. de frequentie, en tenslotte wordt de faseschatting rechtstreeks berekend.

In het geval van M-PSK ($M > 2$) wordt $p(\mathbf{r}|\mathbf{a}, \theta, F, \tau)$ voor grote N_0 benaderd door in de Taylorreeks van $\exp(v)$ termen tot en met v^M te beschouwen. De gezamenlijke NDA schatting van (θ, F, τ) en de afzonderlijke NDA schatting van τ leiden hier tot dezelfde schatting van de timing : $\hat{\tau}_{\text{NDAML}}$ maximaliseert $L_\tau(\hat{\tau})$, die dezelfde uitdrukking aanneemt als hierboven. De

schatting voor F maximaliseert $L_F(\hat{F}) = \sum_k z_k^M(\hat{\tau}_{\text{NDAML}}) \exp(-j2\pi M\hat{F}kT)$ terwijl de schatting

voor θ gegeven is door $\hat{\theta}_{\text{NDAML}} = \frac{1}{M} \arg(L_F(\hat{F}_{\text{NDAML}}))$

Performantie

Laten we bij afwezigheid van ruis de te maximaliseren functies onderzoeken die aanleiding geven tot de schattingen van timing, frequentie en fase.

De schatting van de timing resulteert uit het maximaliseren van

$$\begin{aligned} L_\tau(\hat{\tau}) &= \sum_{k=-K}^K \left| \int r(t) p^*(t - kT - \hat{\tau}) dt \right|^2 \\ &\stackrel{(i)}{\cong} \sum_{k=-K}^K \left| \int \sum_{m=-\infty}^{+\infty} a_m p(t - mT - \tau) \exp(j2\pi FkT + \theta) p^*(t - kT - \hat{\tau}) dt \right|^2 \\ &= \sum_{k=-K}^K \left| \sum_{m=-\infty}^{+\infty} a_m g(kT - mT + \varepsilon_\tau) \right|^2 \end{aligned}$$

De benadering (i) volgt uit het weglaten van de ruis, de trage variatie van $\exp(j2\pi Ft)$, en we hebben gemakshalve $E_s=1$ ondersteld. Hieruit volgt

$$E[L_\tau(\hat{\tau})] = \sum_{k=-K}^K \sum_{m=-\infty}^{+\infty} |g(kT - mT + \varepsilon_\tau)|^2 = (2K+1) \sum_{m=-\infty}^{+\infty} |g(mT + \varepsilon_\tau)|^2$$

De resulterende functie $E[L_\tau(\hat{\tau})]$ is periodisch in de schattingsfout ε_τ , : er treedt een dubbelzinnigheid op van een veelvoud van T in de schatting van de timing. Dit komt omdat in het geval van random datasymbolen de uitgezonden signalen $s(t; \mathbf{a}, \theta, F, \tau)$ en $s(t-T; \mathbf{a}, \theta, F, \tau) = s(t; \mathbf{a}, \theta, F, \tau+T)$ dezelfde statistische eigenschappen hebben. Wanneer de bandbreedte B van $p(t)$ (en dus ook van $g(t)$) voldoet aan $1/(2T) < B < 1/T$, voldoet de bandbreedte $2B$ van $|g(t)|^2$ aan $1/T < 2B < 2/T$. In dit geval bevat de Fourierreeks van $E[L_\tau(\hat{\tau})]$ enkel de constante term en de termen met frequenties $1/T$ en $-1/T$. Het verloop van $E[L_\tau(\hat{\tau})]$ is afgebeeld in Fig. 2. De dubbelzinnigheid kan opgelost worden door een trainingsequentie toe te voegen aan de te versturen data. Het NDA algoritme voor de schatting van τ maakt geen gebruik van de a priori kennis m.b.t de trainingsequentie. Deze kennis wordt pas achteraf aangewend om de dubbelzinnigheid op te lossen nadat de timing (op een veelvoud van T na) is geschat.

Laten we de schattingen van de frequentie en de fase beschouwen in het geval van M-PSK, waarbij we aannemen dat de timing a priori gekend is. Verder onderstellen we dat er geen ISI optreedt aan de uitgang van het matched filter, en nemen we $E_s = 1$. De schatting van de frequentie maximaliseert $|L_F(\hat{F})|$. Op basis van deze onderstellingen geldt dat $z_k(\tau) = a_k \exp(j(2\pi FkT + \theta)) + n_k$, zodat $E[z_k^M(\tau)] = \exp(j(2\pi MFkT + M\theta))$. Hieruit volgt

$$E[L_F(\hat{F})] = \frac{\sin(\pi M \varepsilon_F (2K+1)T)}{\sin(\pi M \varepsilon_F T)} \exp(jM\theta), \quad \text{zodat} \quad |E[L_F(\hat{F})]| = \left| \frac{\sin(\pi M \varepsilon_F (2K+1)T)}{\sin(\pi M \varepsilon_F T)} \right|. \quad \text{Het}$$

verloop van $|E[L(\hat{F})]|$ is geschetst in Fig. 3. Merk op dat de afstand tussen de posities van het globaal maximum en de naburige secundaire maxima met een factor M gereduceerd is t.o.v. het geval waarbij de frequentieschatting gebeurt op basis van een trainingsequentie. In de onderstelling dat de frequentie en de timing a priori gekend zijn, correspondeert de schatting van de fase met het maximaliseren van $L_\theta(\hat{\theta}) = \text{Re}[\exp(-jM\hat{\theta})L_F(F)]$. Uit het bovenstaande volgt $E[L_\theta(\hat{\theta})] = (2K+1) \cos(M\varepsilon_\theta)$. Fig. 4 toont het verloop van $E[L_\theta(\hat{\theta})]$ voor QPSK ($M=4$). Merk op dat er dubbelzinnigheid optreedt : de fase kan maar op een veelvoud van $2\pi/M$ na bepaald worden. Dit is omdat bij random symbolen de uitgezonden signalen $s(t; \mathbf{a}, \theta, F, \tau)$ en $s(t; \mathbf{a} \exp(j2\pi/M), \theta, F, \tau) = s(t; \mathbf{a}, \theta + 2\pi/M, F, \tau)$ dezelfde statistiek hebben, en dus

niet van elkaar kunnen onderscheiden worden. Deze dubbelzinnigheid kan opgelost worden door middel van een trainingsequentie, of door differentiële encoding. Merk op dat in dit geval bij differentiële codering geen differentiële detectie gebeurt, maar wel coherente detectie (op basis van de geschatte fase); dit leidt tot de decisies \hat{a}_{k-1} en \hat{a}_k , waaruit de decisie m.b.t. de differentiële informatie c_k (met $a_k = c_k a_{k-1}$) wordt bekomen als $\hat{c}_k = \hat{a}_k \hat{a}_{k-1}^*$.

De gemiddelde kwadratische schattingsfout bij NDA algoritmes is hoger dan bij DA algoritmes, omdat bij DA schatting de trainingsymbolen gekend zijn. De gemiddelde kwadratische fout bij NDA schatting kan berekend worden aan de hand van een afgebroken Taylorreeks van de te maximaliseren functie (deze berekeningen zijn analoog aan het benaderen van de logaritmische kansfunctie door een afgebroken Taylorreeks). Maakt de schatting \hat{x}_{NDA} van de scalaire parameter x de functie $L_x(\hat{x})$ maximaal, dan geldt, voor \hat{x}

dicht bij de correcte waarde x , $L_x(\hat{x}) \cong L_x(x) + \epsilon_x \left. \frac{dL_x}{d\hat{x}} \right|_{\hat{x}=x} + \frac{1}{2} \epsilon_x^2 \left. \frac{d^2L_x}{d\hat{x}^2} \right|_{\hat{x}=x}$, waarbij

$\epsilon_x = \hat{x} - x$. Hieruit volgt dat de schattingsfout die correspondeert met de NDA schatting

\hat{x}_{NDA} voldoet aan $\hat{x}_{\text{NDA}} - x \cong - \left(\left. \frac{dL_x}{d\hat{x}} \right|_{\hat{x}=x} \right) \left(\left. \frac{d^2L_x}{d\hat{x}^2} \right|_{\hat{x}=x} \right)^{-1} \cong - \left(\left. \frac{dL_x}{d\hat{x}} \right|_{\hat{x}=x} \right) \left(E \left[\left. \frac{d^2L_x}{d\hat{x}^2} \right|_{\hat{x}=x} \right] \right)^{-1}$,

zodat $E[(\hat{x}_{\text{NDA}} - x)^2] \cong E \left[\left(\left. \frac{dL_x}{d\hat{x}} \right|_{\hat{x}=x} \right)^2 \right] \left(E \left[\left. \frac{d^2L_x}{d\hat{x}^2} \right|_{\hat{x}=x} \right] \right)^{-2}$. Alhoewel conceptueel eenvoudig,

kunnen deze berekeningen vrij bewerkelijk zijn, zodat in plaats van analytische berekeningen vaak computersimulaties worden gebruikt om de gemiddelde kwadratische schattingsfout te bepalen.

Op basis van de kansfunctie $p(\mathbf{r}|\theta, F, \tau)$ van de parameters (die wordt bekomen door de gezamenlijke kansfunctie $p(\mathbf{r}|\mathbf{a}, \theta, F, \tau)$ uit te middelen over \mathbf{a}) kan in principe de Cramer-Rao grens bepaald worden, maar wegens de ingewikkelde vorm van $p(\mathbf{r}|\theta, F, \tau)$ is dit veel moeilijker dan bij de transmissie van een trainingsequentie. Om deze reden wordt als benedengrens op de gemiddelde kwadratische schattingsfout vaak gewoon de Cramer-Rao grens voor de transmissie van een (lange) trainingsequentie genomen, alhoewel deze (vooral bij lage E_s/N_0) minder dicht is dan de Cramer-Rao grens voor random data.

Detectie op basis van geschatte parameters

Bij de strategieën 2, 3 en 4 wordt bij de detectie gebruik gemaakt van een schatting van de parameter \mathbf{x} . De nauwkeurigheid van deze schatting heeft een invloed op de betrouwbaarheid van de detectie.

Fig. 5 illustreert het effect van een fasefout ε_θ op de BER bij QPSK. We merken op dat de waarde van E_b/N_0 , nodig om een bepaalde streefwaarde van de BER te bekomen, toeneemt naarmate $|\varepsilon_\theta|$ stijgt. Willen we de vereiste toename van E_b/N_0 t.o.v. het geval van perfecte faseschatting beperken tot een kleine fractie van een dB (bijvoorbeeld 0.1 dB), dan mag $|\varepsilon_\theta|$ slechte enkele graden bedragen : voor een degradatie van ten hoogste 0.1 dB bij $BER = 10^{-2}$ en $BER = 10^{-6}$ mag $|\varepsilon_\theta|$ niet groter zijn dan respectievelijk 3.4 graden en 1.8 graden.

Schattingsalgoritmes die gebruik maken van feedback

Tot nu toe hebben we de situatie beschouwd waarbij op basis van het ontvangen signaal $r(t)$ een schatting wordt gemaakt van een parameter \mathbf{x} , door middel van een zoekprocedure of (indien mogelijk) door rechtstreekse berekening : dit is een 'feedforward' aanpak. Hierbij hebben we stilzwijgend aangenomen dat de parameter \mathbf{x} nagenoeg constant is over het beschouwde observatieinterval. Gaat deze benadering niet op, dan kunnen we het observatieinterval opsplitsen in deelintervallen, waarbinnen de parameter \mathbf{x} wel als constant kan beschouwd worden, en voor elk deelinterval een afzonderlijke schatting maken van \mathbf{x} .

Een alternatief is het maken van een schatting met behulp van 'feedback'. In dit geval wordt de schatting van de parameter \mathbf{x} op geregelde tijdstippen aangepast. De correctie van de schatting op een bepaald tijdstip gebeurt op basis van een indicatie van de schattingsfout op dit tijdstip en op vroegere tijdstippen. We illustreren deze methode aan de hand van een voorbeeld : het schatten van een traag variërende fase bij lineaire digitale modulatie, waarbij de schatting van de fase eenmaal per symboolinterval wordt aangepast.

Het ontvangen signaal $r(t)$ is gegeven door $r(t) = \sqrt{E_s} \sum_k a_k p(t - kT) \exp(j\theta) + w(t)$, waarbij

$\{a_k\}$ en θ respectievelijk een trainingsequentie en de te schatten fase voorstellen. We nemen aan dat er geen ISI optreedt na het matched filter. De logaritmische kansfunctie (op irrelevante constanten na) is gegeven door $\ln p(\mathbf{r} | \theta) = \sum_k \text{Re}[a_k^* z_k \exp(-j\theta)]$. Afleiden naar θ

geeft $\frac{\partial \ln p(\mathbf{r} | \theta)}{\partial \theta} = \sum_k \text{PED}_k(\theta)$, waarbij $\text{PED}_k(\theta) = \text{Im}[a_k^* z_k \exp(-j\theta)]$ ('PED' slaat op 'phase

error detector' : we zullen aantonen dat $\text{PED}_k(\cdot)$ een indicatie geeft van de fasefout). De ML schatting $\hat{\theta}_{ML}$ maakt de afgeleide van de logaritmische kansfunctie nul. Opeenvolgende benaderingen van $\hat{\theta}_{ML}$ kunnen als volgt worden bekomen : $\hat{\theta}_{ML}^{(i+1)} = \hat{\theta}_{ML}^{(i)} + \alpha \sum_k \text{PED}_k(\hat{\theta}_{ML}^{(i)})$,

waarbij de parameter $\alpha (>0)$ de convergentie beïnvloedt (kleine α leidt tot trage convergentie; te grote α leidt tot divergentie). Merk op dat de sommatie (over k) loopt over het volledige observatieinterval. Beschouw nu de situatie waarbij we per symboolinterval een schatting willen maken op basis van de tot dan gemaakte observatie. We noemen $\hat{\theta}_k$ de schatting die

correspondeert met het k -de symboolinterval ($k = 0, 1, 2, \dots$). Naar analogie met de iteraties voor het bepalen van $\hat{\theta}_{ML}$ stellen we :

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \alpha \sum_{m=0}^k f_{k-m} \text{PED}_m(\hat{\theta}_m) \quad \text{voor } k = 0, 1, 2, \dots, \text{ en waarbij } \hat{\theta}_0 \text{ willekeurig in } (-\pi, \pi)$$

wordt gekozen. Fig. 6 toont een structuur die de schattingen $\hat{\theta}_k$ berekent. Hierbij is $F(z)$ de z -getransformeerde van f_k : $F(z) = \sum_{k=0}^{\infty} f_k z^{-k}$. Met behulp van een tabel wordt $\exp(-j\hat{\theta}_k)$

berekend uit $\hat{\theta}_k$. De transferfunctie z^{-1} stelt een tijdsvertraging over een symboolinterval T voor. Deze structuur is verwant met een fasevergrendelingslus (phase-locked loop, PLL).

Bemerk de volgende verschillen tussen de iteraties voor het bepalen van $\hat{\theta}_{ML}$ en de berekening van $\hat{\theta}_k$:

- Voor het berekenen van $\hat{\theta}_{k+1}$ maken we enkel gebruik van gegevens uit vorige symboolintervallen (met indices $k, k-1, \dots$)
- De functie $\text{PED}_m(\cdot)$ wordt slechts éénmaal berekend, nl. tijdens het symboolinterval met index m , en voor het argument $\hat{\theta}_m$.
- De waarde $\text{PED}_m(\hat{\theta}_m)$ heeft effect op de schattingen in de intervallen met index $m+1, m+2, m+3, \dots$. De bijdrage tot $\hat{\theta}_{k+1}$ wordt gewogen met een coëfficiënt f_{k-m} . Dit laat toe om bijvoorbeeld meer gewicht te hechten aan het recent verleden dan aan het verre verleden.

Laten we aannemen dat de fase traag fluctueert (een kleine frequentiefout kan in de fluctuerende fase worden opgenomen). Dit wordt in rekening gebracht door in de definitie van $r(t)$ de constante fase θ te vervangen door θ_k , de fase die correspondeert met het k -de symboolinterval. Laten we $\text{PED}_m(\hat{\theta}_m)$ ontbinden als de som van het gemiddelde en een statistische fluctuatie met gemiddelde nul : $\text{PED}_m(\hat{\theta}_m) = E[\text{PED}_m(\hat{\theta}_m)] + n_m(\hat{\theta}_m)$. Uit de vorm van $r(t)$ volgt dat $E[\text{PED}_m(\hat{\theta}_m)] = -s(\epsilon_m)$, waarbij $\epsilon_m = \hat{\theta}_m - \theta_m$ de fasefout en $s(\epsilon_m) = \sqrt{E_s} \sin(\epsilon_m)$ de fasefasedetector karakteristiek (ook nog 's-curve' genoemd) voorstelt. Hieruit blijkt dat we $\text{PED}_m(\hat{\theta}_m)$ kunnen beschouwen als een door de fluctuatie $n_m(\hat{\theta}_m)$ verstoorte indicatie van de fasefout ϵ_m ; 'PED' verwijst dan ook naar 'phase error detector'.

Laten we het verloop van de fasefout ϵ_k onderzoeken, in de onderstelling dat θ_k heel traag varieert, en de fluctuatie $n_k(\hat{\theta}_k)$ klein is. Uit de vergelijking voor $\hat{\theta}_{k+1}$ volgt een vergelijking

$$\text{voor de fasefout : } \epsilon_{k+1} = \epsilon_k + \underbrace{\theta_k - \theta_{k+1}}_{\text{klein}} - \alpha \sum_{m=0}^k f_{k-m} \left(s(\epsilon_m) - \underbrace{n_m(\hat{\theta}_m)}_{\text{klein}} \right).$$

($\theta_k - \theta_{k+1}$) en $N_m(\hat{\theta}_m)$ verwaarlozen, dan convergeert ϵ_k van de beginwaarde ϵ_0 naar de regimewaarde nul ($s(0) = 0$). In werkelijkheid vertoont de fasefout in regime een kleine fluctuatie, die mede bepaald is door de incrementen ($\theta_{k+1} - \theta_k$) en de stoortermen $n_k(\hat{\theta}_k)$. De evolutie van de (eventueel grote) beginwaarde ϵ_0 naar de regimetoestand (waar de fasefout kleine waarden aanneemt) wordt de *acquisitiemode* genoemd. Voor kleine fasefouten gelden de volgende benaderingen :

$s(\varepsilon_m) \cong \dot{s}(0)\varepsilon_m$ (waarbij $\dot{s}(0) = \sqrt{E_s}$ de helling van de fase-detector karakteristiek voorstelt) en $n_m(\hat{\theta}_m) \cong n_m(\theta_m)$, zodat het verloop van de fasefout in regime bepaald wordt door de volgende vergelijking :

$$\varepsilon_{k+1} - \varepsilon_k + \alpha \dot{s}(0) \sum_{m=0}^k f_{k-m} \varepsilon_m = \theta_k - \theta_{k+1} + \alpha \sum_{m=0}^k f_{k-m} n_m(\theta_m).$$

Stellen we de z-getransformeerden van ε_k , θ_k , $\hat{\theta}_k$ en $n_k(\theta_k)$ voor door $E(z)$, $\Theta(z)$, $\hat{\Theta}(z)$ en $N(z)$, dan bekommen we

$$E(z) = H(z) \frac{N(z)}{\dot{s}(0)} - (1 - H(z))\Theta(z) \text{ en } \hat{\Theta}(z) = H(z) \left(\Theta(z) + \frac{N(z)}{\dot{s}(0)} \right), \text{ waarbij}$$

$$H(z) = \frac{\alpha \dot{s}(0) F(z)}{z - 1 + \alpha \dot{s}(0) F(z)} \text{ en } 1 - H(z) = \frac{z - 1}{z - 1 + \alpha \dot{s}(0) F(z)}$$

de transferfuncties voorstellen van respectievelijk een laagdoorlaatfilter ($H(1) = 1$) en een hoogdoorlaatfilter. In regime 'volgt' de schatting $\hat{\theta}_k$ dus (een gefilterde versie van) θ_k ; dit noemt men de *tracking mode*. Stellen we de bandbreedte van het laagdoorlaatfilter voor door B_L , dan bestaat de fasefout uit de frequentiecomponenten van $n_k(\theta_k)$ die binnen de band $(-B_L, B_L)$ zijn gelegen, en uit de frequentiecomponenten van θ_k die buiten de band $(-B_L, B_L)$ zijn gelegen. Hieruit volgt dat een geschikte waarde van B_L een compromis is : het verkleinen (vergroten) van B_L verkleint (vergroot) de bijdrage van $n_k(\theta_k)$, maar vergroot (verkleint) de bijdrage van de fluctuatie van θ_k . De waarde van B_L heeft ook een effect op de acquisitiemodus : hoe kleiner B_L , hoe langer de acquisitiemodus.

Fig. 7 toont het verloop van de fasefout, in de onderstelling dat $FT = -10^{-4}$, $F(z) = 1$ (zodat $f_k = \delta_k$) en de initiële fasefout 45 graden bedraagt. Wegens de van nul verschillende frequentieoffset F evolueert de fasefout bij afwezigheid van ruis naar een van nul verschillende statische regimewaarde, die des te groter is naarmate $B_L T$ kleiner is (ongeveer 0.9 graden en 9 graden voor respectievelijk $B_L T = 10^{-2}$ en $B_L T = 10^{-3}$). De duur van de acquisitie neemt toe wanneer $B_L T$ kleiner wordt (ongeveer 125 en 1250 symboolintervallen voor respectievelijk $B_L T = 10^{-2}$ en $B_L T = 10^{-3}$). Bij aanwezigheid van ruis vertoont de fasefout statistische fluctuaties, die des te groter zijn naarmate $B_L T$ toeneemt (de spreiding bedraagt ongeveer 4.5 graden en 1.4 graden voor respectievelijk $B_L T = 10^{-2}$ en $B_L T = 10^{-3}$).

De hierboven besproken feedbackstructuur heeft betrekking op een DA faseschatting (de datasymbolen van de trainingsequentie zijn a priori gekend). Hetzelfde principe kan toegepast worden op DD en NDA faseschatting, en op DA, DD en NDA schatting van de timing. (Oefening : teken een blokschema van de resulterende structuren, uitgaande van de functies die door de corresponderende feedforward-schattingen worden gemaximaliseerd).

Enkele voordelen van een feedbackstructuur zijn :

- Een fluctuerende parameter kan 'gevolgd' worden (zodat een kleine frequentiefout kan opgenomen worden in de fase, en de frequentie dus niet afzonderlijk moet geschat worden)
- Het schatten van de parameter vereist geen zoekprocedure, en is dus minder rekenintensief.

Een nadeel van een feedbackstructuur is de acquisitiemodus, gedurende dewelke de schatting niet nauwkeurig is, en de decisies m.b.t. de datasymbolen bijgevolg weinig betrouwbaar zijn. Het is dus aanbevolen tijdens de acquisitie een trainingsequentie te versturen, waarvan de lengte correspondeert met de duur van de acquisitiemodus.

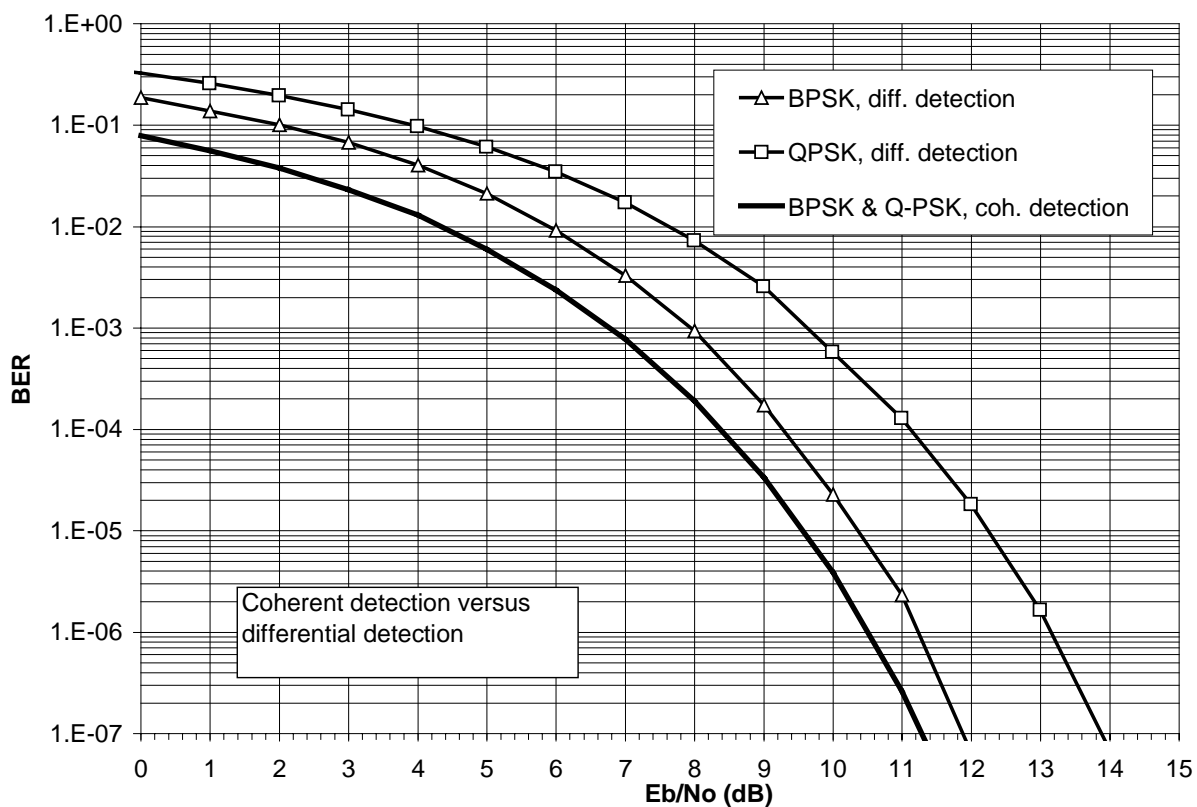


Fig. 1 : vergelijking van coherente detectie en differentiële detectie bij BPSK en QPSK

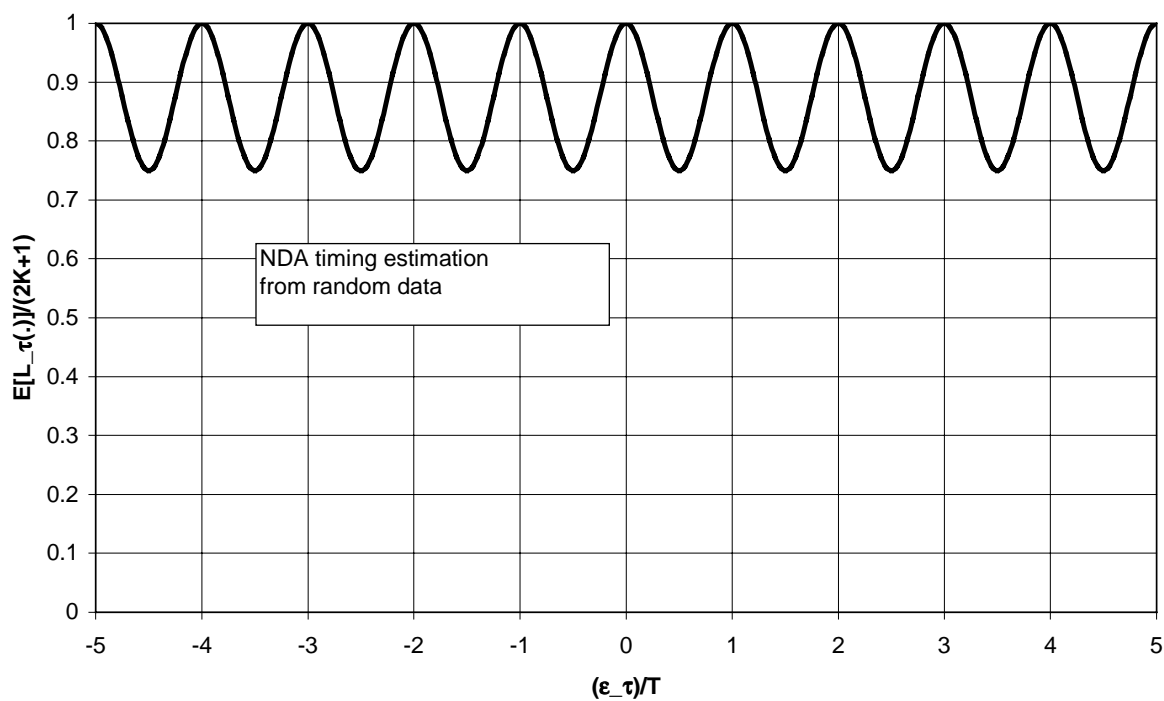


Fig. 2 : te maximaliseren functie bij NDA schatting van de timing

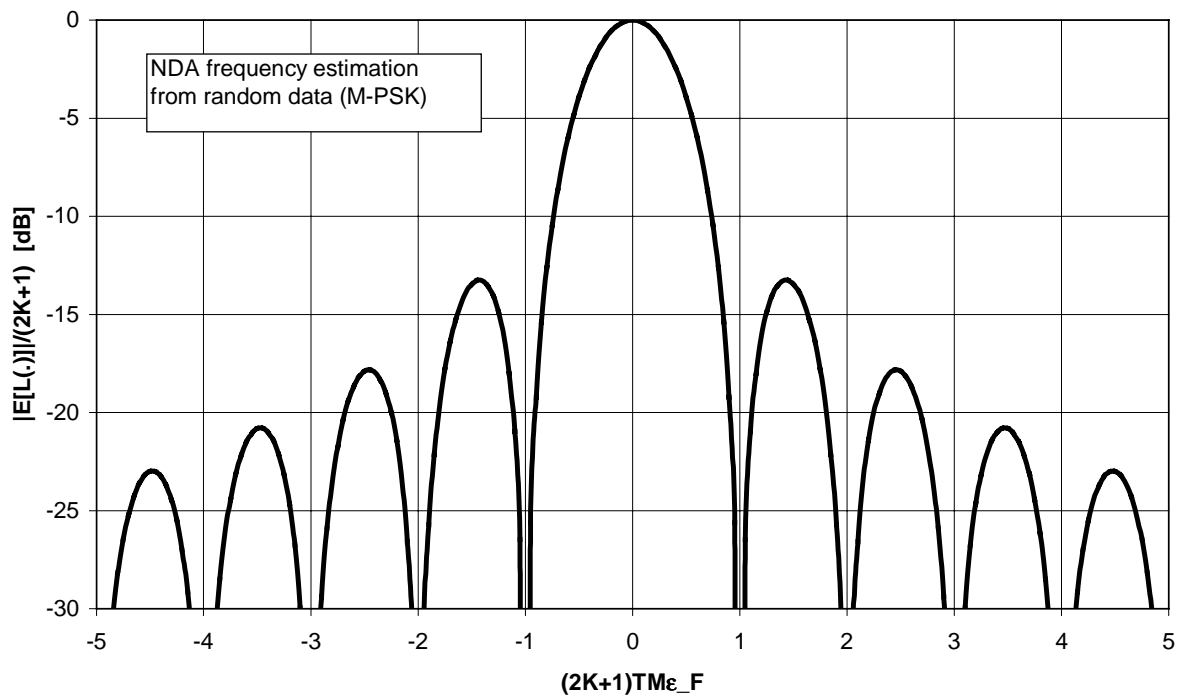


Fig. 3 : te maximaliseren functie bij NDA schatting van de frequentie

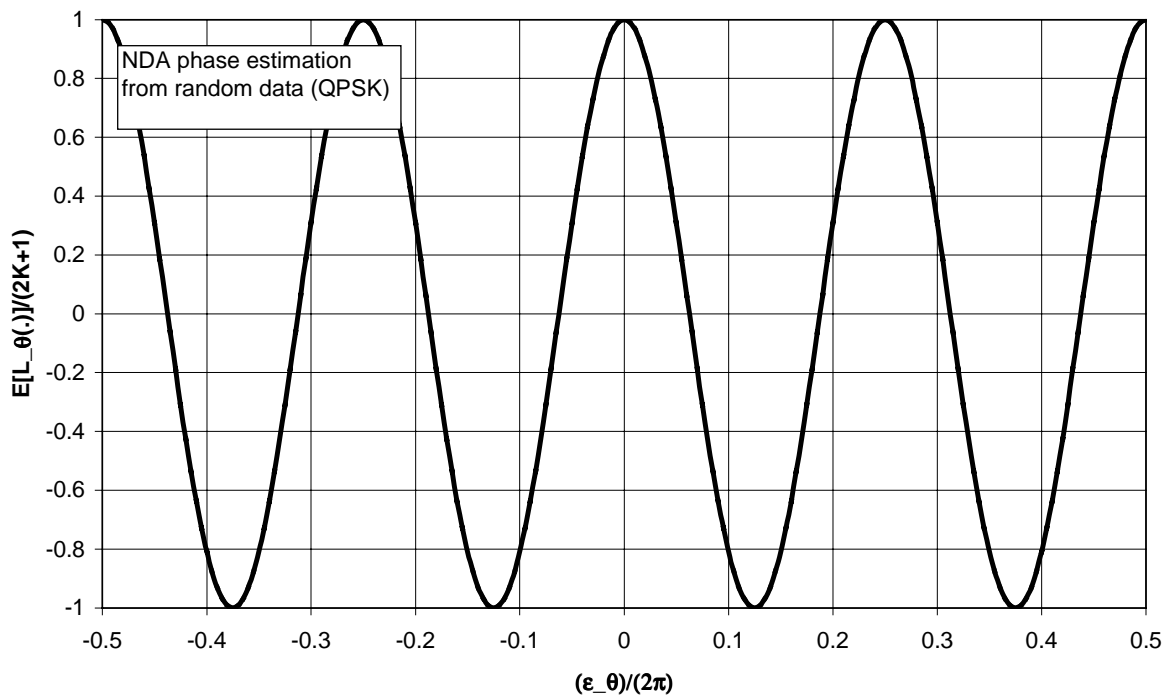


Fig. 4 : te maximaliseren functie bij NDA schatting van de fase

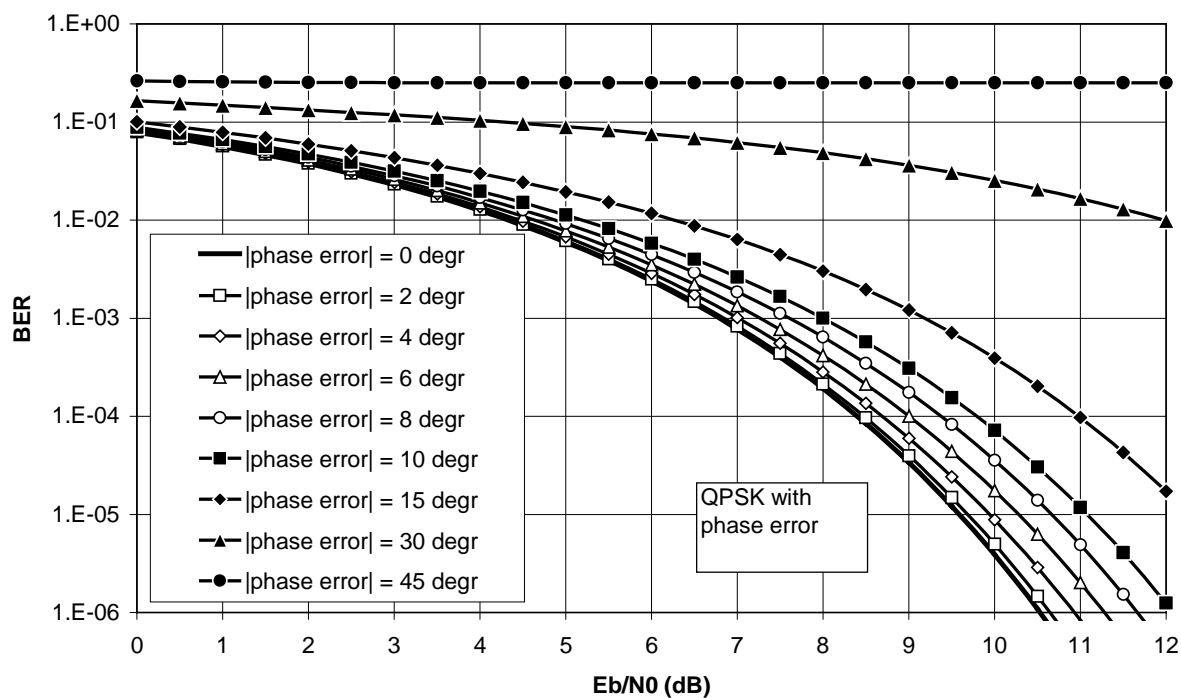


Fig. 5 : effect van fasefout op BER (coherente detectie van QPSK)

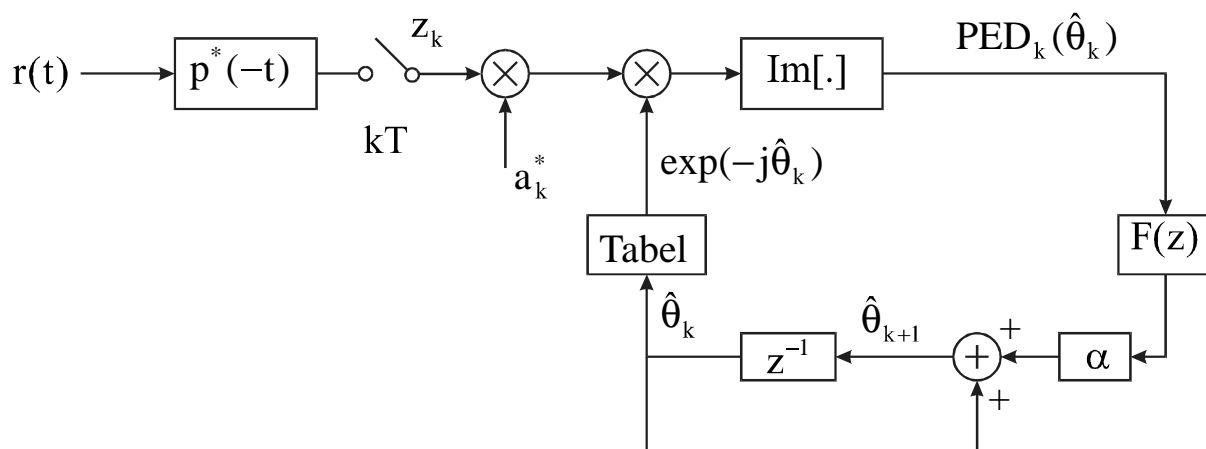


Fig. 6 : feedback-structuur voor faseschatting

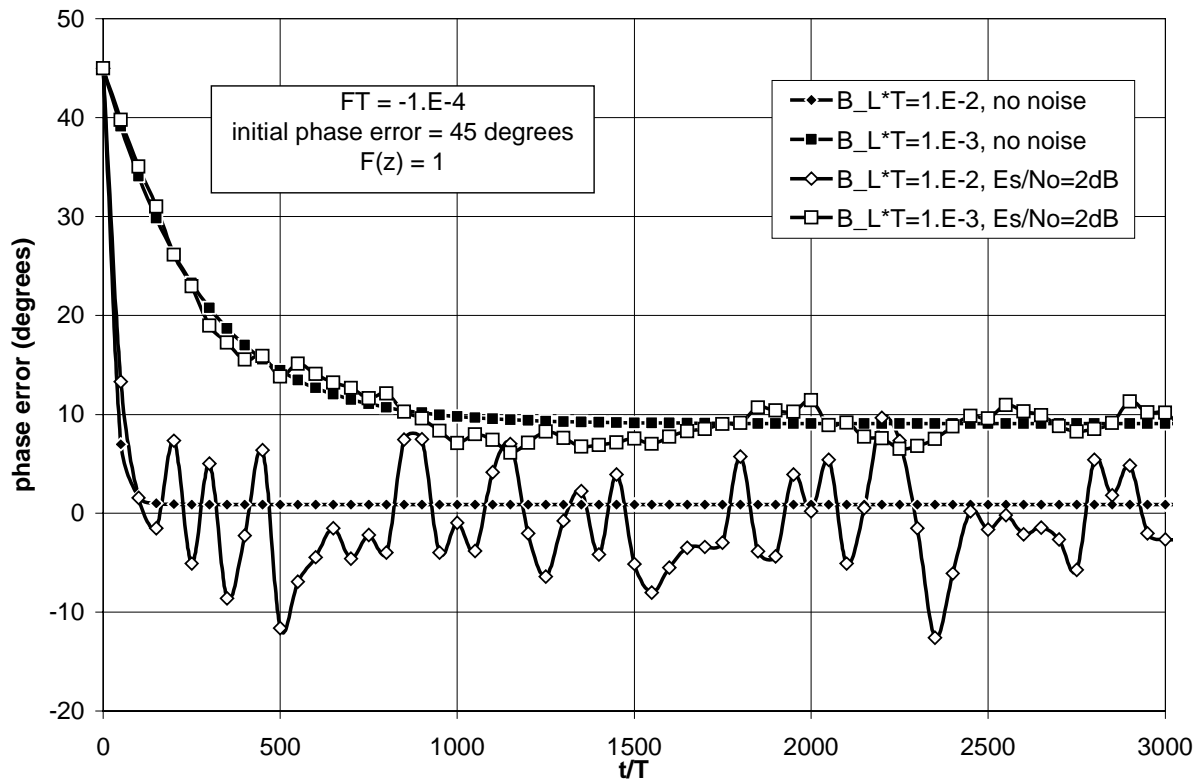


Fig. 7 : verloop van de fasefout bij schatting met behulp van feedback.