

SCUOLA DI INGEGNERIA

Sede di Forlì

Corso di Laurea in
INGEGNERIA AEROSPAZIALE

Classe L - 9

Elaborato finale di Laurea in
ING-IND/05: Fondamenti di Meccanica Orbitale

**IMPLEMENTAZIONE E CARATTERIZZAZIONE
DI PLL DIGITALI PER DOPPLER TRACKING DI
SONDE DEEP SPACE**

CANDIDATO

Stefano Palmiotto

RELATORE

Paolo Tortora

CORRELATORE

Andrea Togni

Anno Accademico 2021/2022

Indice

1	Introduzione	1
1.1	Determinazione d'orbita statistica	1
1.1.1	Il processo di determinazione d'orbita	1
1.1.2	Tracking e acquisizione dati	9
1.1.3	Osservabili	11
1.1.4	Fonti di errore	15
1.2	Digital Signal Processing	19
1.2.1	Rappresentazione dei segnali	19
1.2.2	Segnali analogici, segnali digitali e campionamento	23
2	Phase-Locked Loops	26
2.1	Caratteristiche principali	26
2.2	Phase Detector	28
2.3	Lowpass Filter	29
2.4	Voltage-Controlled Oscillator	29
2.5	Analisi lineare	30
2.6	Digital Phase-Locked Loops	32
2.7	Implementazione in Matlab e Simulink	33
2.8	Verifica del modello	38
3	La missione Voyager	41
3.1	Storia della missione Voyager	41
3.2	Architettura e sottosistemi della sonda	42
3.3	Gli esperimenti nel sistema di Nettuno	43
3.4	Analisi del segnale con il PLL	44

Elenco delle figure

1.1	Problema dei due corpi	2
1.2	Coordinate cartesiane, cilindriche e sferiche	4
1.3	Residui	7
1.4	Determinazione orbitale	8
1.5	Bande radio	9
1.6	Sequenza di acquisizione del downlink	10
1.7	Range-rate	13
1.8	Distribuzione normale	18
1.9	AWGN Spectrum	18
1.10	GPS	19
1.11	Rappresentazione temporale	20
1.12	Spettro	21
1.13	Simmetria hermitiana	22
1.14	Formula di Eulero	23
1.15	Conversione di un segnale dal formato analogico a quello digitale	24
1.16	Aliasing di un segnale sinusoidale sotto-campionato	25
1.17	Aliasing nello spettro del segnale	25
2.1	Schema del PLL	26
2.2	Definizioni di Loop Bandwidth	27
2.3	PLL unlocked state	32
2.4	Unwrap	33
2.5	Unwrap anomalo	34
2.6	Simulink PLL Model	34
2.7	Lowpass Filter	35
2.8	Numerically-Controlled Oscillator	36
2.9	Periodogramma	37
2.10	Errore di fase a regime PLL II	38
2.11	Errore di fase a regime PLL III	39
2.12	Limite inferiore di Cramer-Rao	40
2.13	Limite inferiore di Cramer-Rao per banda di loop ottimale	40
3.1	Voyager 2	42
3.2	Neptune radio occultation	43
3.3	Segnale in banda X	45
3.4	Stima di frequenza in banda X	46
3.5	Frequenza di antenna in banda X	46
3.6	Stima di frequenza in banda S	47
3.7	Frequenza di antenna in banda S	47

3.8	Confronto delle stime di frequenza	48
-----	--	----

Elenco delle tabelle

2.1	Errore di fase a regime	31
2.2	Coefficienti del JPL	38

*A Piero e tutte le persone che hanno sempre mantenuto viva la mia curiosità di
conoscere il mondo*

Alla mia famiglia, in particolare a Giovanni che sarebbe stato fiero di questo traguardo

A tutte le persone che hanno reso preziosi alcuni momenti del mio percorso

Sommario

I Phase-Locked Loops sono circuiti ancora oggi utilizzati per la generazione di segnali coerenti in frequenza e in fase con i segnali in ingresso, motivo per cui sono uno degli strumenti della radio scienza per la ricostruzione dei segnali scambiati con le sonde e nascosti dal rumore accumulato nel tragitto che separa le sonde stesse dalle stazioni di tracking a terra.

Questa tesi illustra l'implementazione di un PLL digitale linearizzato in Matlab e Simulink in una nuova veste rispetto al modello implementato durante l'attività di tirocinio curricolare, al fine di migliorarne le prestazioni per bassi carrier-to-noise density ratios.

Il capitolo 1 si compone di due parti: la prima introduce all'ambito nel quale si inserisce il lavoro proposto, ossia la determinazione d'orbita; la seconda illustra i fondamenti della teoria dei segnali.

Il capitolo 2 è incentrato sull'analisi dei Phase-locked Loops, partendo da un'introduzione teorica e approdando all'implementazione di un modello in Simulink.

Il capitolo 3, infine, mostra i risultati dell'applicazione del modello implementato in Simulink nell'analisi dei segnali di una missione realmente svolta.

Capitolo 1

Introduzione

1.1 Determinazione d'orbita statistica

1.1.1 Il processo di determinazione d'orbita

Il primo caso di determinazione d'orbita risale al lavoro di Johannes Kepler (1571-1630), il quale, a partire dai dati delle osservazioni astronomiche di Tycho Brahe (1546-1601) formulò le seguenti tre leggi:

1. i pianeti descrivono orbite ellittiche, di cui il Sole occupa uno dei due fuochi
2. il segmento che congiunge il centro del Sole e il centro del pianeta descrive aree uguali in tempi uguali
3. il rapporto tra il cubo del semiasse maggiore dell'orbita e il quadrato del periodo orbitale è lo stesso per tutti i pianeti

Fu poi Isaac Newton (1642-1726) a formulare i fondamenti matematici della meccanica e della gravitazione dai quali derivare tali leggi e a pubblicarli nei *Principia* su esortazione di Edmond Halley (1656-1742). I cardini della trattazione di Newton sono i concetti di massa puntuale di un corpo, cioè concentrata in un punto detto centro di massa, di sistema di riferimento inerziale, di forza e di spazio e tempo assoluti. Le leggi del moto di Newton sono alla base della meccanica orbitale e ancora oggi forniscono una descrizione sufficientemente accurata del moto di satelliti naturali e artificiali. Una maggiore accuratezza è data dalla teoria della relatività di Albert Einstein (1879-1955), la quale è anche in grado di spiegare le piccole discrepanze tra le osservazioni e le predizioni del modello newtoniano [1]. Il punto di partenza è la definizione di un opportuno sistema di riferimento inerziale nel quale valga il secondo principio della dinamica:

$$\mathbf{F} = m\mathbf{a} = m\ddot{\mathbf{R}} \quad (1.1)$$

essendo \mathbf{F} la forza, m la massa inerziale, cioè la proprietà intrinseca di un corpo di opporsi ad una variazione del suo stato di quiete o di moto rettilineo uniforme, e $\mathbf{a} = \ddot{\mathbf{R}}$ l'accelerazione, definita come la derivata seconda rispetto al tempo della posizione \mathbf{R} , quest'ultima definita rispetto all'origine del sistema di riferimento che scegliamo di adottare. La forza di attrazione gravitazionale secondo la legge di Newton è così definita:

$$\mathbf{F}_{g,12} = -\frac{Gm_1m_2}{r^3}\mathbf{r} \quad (1.2)$$

dove $\mathbf{F}_{g,12}$ è l'attrazione gravitazionale esercitata dal corpo di massa m_1 su quello di massa m_2 , $G \simeq 6.67 \times 10^{-11} \text{ m}^3\text{kg}^{-1}\text{s}^{-2}$ è la costante di gravitazione universale, misurata per la prima volta da Henry Cavendish (1731-1810), e \mathbf{r} è il vettore che indica la posizione di m_2 relativa a m_1 . Occorre prestare attenzione al fatto che quelle che compaiono nell'Eq. (1.2) non sono masse inerziali, ma gravitazionali, ossia la proprietà che consente ad un corpo di essere sorgente e di subire l'influenza di un campo gravitazionale. Grazie al postulato di equivalenza di Einstein, tuttavia, possiamo considerare la massa inerziale e la massa gravitazionale di un corpo coincidenti.

Facendo riferimento alla Fig. (1.1), siano M e m due masse costituenti un sistema isolato, ciascuna delle quali soggetta all'attrazione gravitazionale dell'altra.

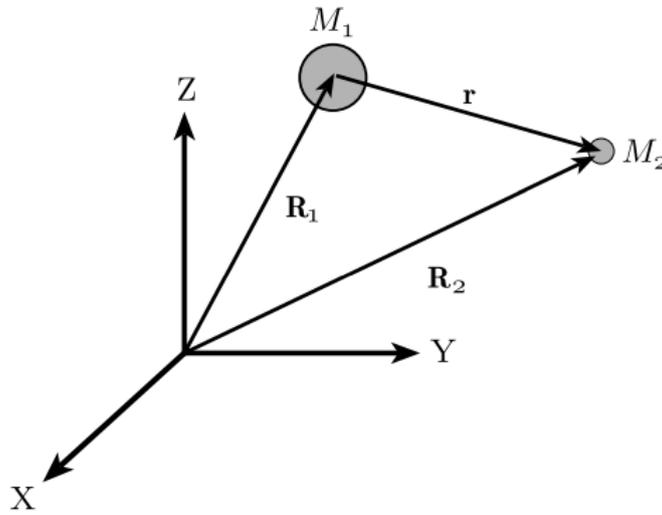


Figura 1.1: Problema dei due corpi. Il sistema di riferimento $X - Y - Z$ è inerziale [1]

Applicando le Eq. (1.1) e (1.2):

$$M_1\ddot{\mathbf{R}}_1 = \frac{GM_1M_2}{r^3}\mathbf{r} \quad (1.3)$$

$$M_2\ddot{\mathbf{R}}_2 = -\frac{GM_2M_1}{r^3}\mathbf{r} \quad (1.4)$$

Osserviamo che le due forze sono opposte, in accordo al principio di azione e reazione e all'ipotesi di sistema isolato. Sottraendo la (1.3) alla (1.4) si giunge alla seguente equazione:

$$\ddot{\mathbf{r}} = -\frac{G(M_1 + M_2)}{r^3}\mathbf{r} = -\frac{\mu}{r^3}\mathbf{r} \quad (1.5)$$

cioè l'accelerazione del corpo di massa M_2 relativa al corpo di massa M_1 . Il modello appena ricavato è noto come problema dei due corpi. Nonostante faccia uso di numerose semplificazioni, tra cui l'aver considerato il contributo delle sole forze gravitazionali e che queste siano centrali, ossia applicate nei centri di massa, esso costituisce un punto di

partenza per l'interpretazione di rappresentazioni più realistiche del moto orbitale, come quelle che tengono conto delle forze di perturbazione:

$$\ddot{\mathbf{r}} = -\frac{\mu}{r^3}\mathbf{r} + \mathbf{f} \quad (1.6)$$

Il vettore \mathbf{f} è l'accelerazione dovuta alla risultante delle forze di perturbazione, equivalente alla somma di diversi contributi [1]:

$$\mathbf{f} = \mathbf{f}_{NS} + \mathbf{f}_{3B} + \mathbf{f}_g + \mathbf{f}_{drag} + \mathbf{f}_{SRP} + \mathbf{f}_{other} \quad (1.7)$$

dove:

- *NS* sta per *Non Spherical*, cioè le perturbazioni gravitazionali dovute alla non-sfericità del corpo intorno al quale orbita il satellite
- *3B* significa *Third Body*, ossia la perturbazione gravitazionale indotta dalla presenza di un terzo corpo. Un esempio è l'interazione reciproca tra il Sole, La Terra e la Luna.
- *g* indica il contributo della relatività generale
- *drag* è la perturbazione non gravitazionale dovuta alla resistenza aerodinamica che interessa i satelliti in orbita bassa intorno a pianeti dotati di un'atmosfera
- *SRP* è la sigla di *Solar Radiation Pressure*, quindi denota la perturbazione non gravitazionale dei fotoni che impattano i corpi in orbita
- il pedice *other* indica altre perturbazioni differenti da quelle precedentemente elencate, come quelle di natura termica

L'Eq. (1.6) è dunque una delle possibili equazioni che descrivono la dinamica orbitale del satellite. Si tratta di un'equazione differenziale ordinaria vettoriale del secondo ordine non lineare, equivalente ad un sistema di 3 equazioni scalari. Le funzioni incognite sono le leggi temporali delle tre componenti del vettore posizione \mathbf{r} del satellite, espresse solitamente in coordinate cartesiane o in coordinate polari sferiche:

$$\mathbf{r}(t) = \begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix} \quad (\text{coordinate cartesiane}) \quad (1.8)$$

$$\mathbf{r}(t) = \begin{bmatrix} r(t) \\ \delta(t) \\ \alpha(t) \end{bmatrix} \quad (\text{coordinate polari sferiche}) \quad (1.9)$$

dove δ è la declinazione e α è l'ascensione retta, ossia i due angoli che definiscono l'orientazione del vettore \mathbf{r} nel sistema di riferimento, come illustrato in Fig. (1.2).

Le condizioni iniziali dell'Eq. (1.6) sono:

$$\mathbf{r}(t_o) = \mathbf{r}_o \quad (1.10)$$

$$\dot{\mathbf{r}}(t_o) = \mathbf{v}(t_o) = \mathbf{v}_o \quad (1.11)$$

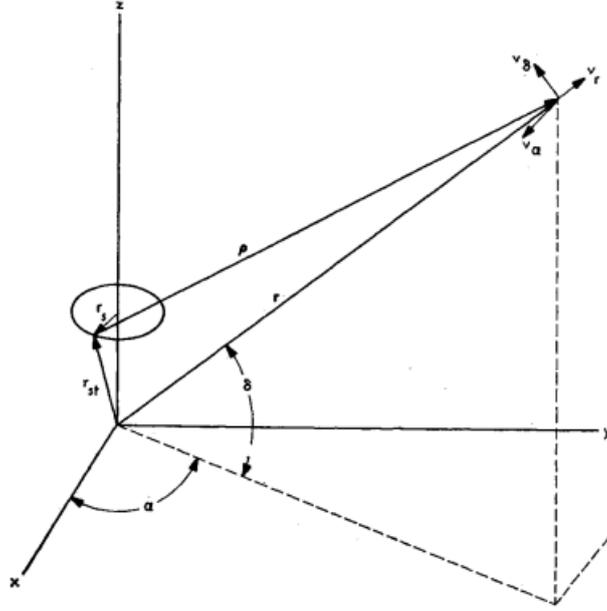


Figura 1.2: Coordinate cartesiane (x, y, z) e coordinate polari sferiche (r, δ, α) [2]

Attraverso una semplice manipolazione, è possibile convertire l'Eq. (1.6) in un sistema di 2 equazioni differenziali ordinarie vettoriali del primo ordine, equivalente ad un sistema di 6 equazioni scalari:

$$\begin{cases} \dot{\mathbf{r}} = \mathbf{v} \\ \dot{\mathbf{v}} = -\frac{\mu}{r^3}\mathbf{r} + \mathbf{f} \end{cases} \quad (1.12)$$

In questo caso le funzioni incognite sono le tre componenti del vettore posizione e le tre componenti del vettore velocità \mathbf{v} , per un totale di sei incognite scalari. Le condizioni iniziali sono sempre quelle indicate in (1.10) e (1.11).

Definiamo il vettore $\boldsymbol{\xi} = \boldsymbol{\xi}(t)$, le cui componenti sono le 6 incognite scalari del sistema (1.12):

$$\boldsymbol{\xi}(t) = \begin{bmatrix} \mathbf{r}(t) \\ \mathbf{v}(t) \end{bmatrix} \quad (1.13)$$

Il sistema (1.12) può essere così riformulato:

$$\dot{\boldsymbol{\xi}} = f(\boldsymbol{\xi}, t) \quad (1.14)$$

con condizioni iniziali $\boldsymbol{\xi}(t_0) = \boldsymbol{\xi}_0$. Il vettore $\boldsymbol{\xi}$ è noto come *vettore di stato* e l'Eq. (1.14) è quindi detta *equazione di stato*. Le componenti di posizione e velocità non sono gli unici parametri che definiscono lo stato del satellite: in alcuni casi si aggiungono ulteriori parametri costanti incogniti che compaiono nelle equazioni della dinamica o nei modelli dei sistemi di misurazione. Sia $\boldsymbol{\alpha}$ il vettore $q \times 1$ le cui componenti sono le suddette costanti. Allora è possibile ampliare il vettore di stato fino a $n = 6 + q$ componenti:

$$\mathbf{X}(t) = \begin{bmatrix} \boldsymbol{\xi}(t) \\ \boldsymbol{\alpha} \end{bmatrix} \quad (1.15)$$

essendo \mathbf{X} il *vettore di stato ampliato*. L'equazione di stato assume, dunque, la seguente forma:

$$\dot{\mathbf{X}}(t) = F(\mathbf{X}, t) = \begin{bmatrix} f(\boldsymbol{\xi}, t) \\ \mathbf{0}_{q \times 1} \end{bmatrix} \quad (1.16)$$

con condizioni iniziali $\mathbf{X}(t_o) = \mathbf{X}_o$. Poiché in questo caso si hanno n equazioni in n incognite, nell'ipotesi puramente ideale di assenza di incertezze nelle misurazioni e nei modelli matematici, basta una singola osservazione per ottenere le condizioni iniziali, dopodiché si procede con l'integrazione delle equazioni differenziali per determinare lo stato del satellite in un qualunque istante di tempo $t \geq t_o$. In merito alle osservazioni, generalmente non è possibile osservare direttamente le variabili di stato, cioè le componenti del vettore \mathbf{X} , tuttavia i dati ottenuti dalle osservazioni, che per questo sono anche detti *osservabili*, sono esprimibili in funzione delle suddette variabili:

$$\mathbf{Y} = G(\mathbf{X}, t) \quad (1.17)$$

dove \mathbf{Y} è il vettore contenente le osservabili e G è una funzione generalmente non lineare.

Nella realtà, tuttavia, lo stato iniziale \mathbf{X}_o del satellite non è mai noto con esattezza e alcune costanti fisiche che compaiono nelle equazioni differenziali del moto sono note con un certo grado di approssimazione, oltre ad inevitabili errori nella formulazione dei modelli matematici. L'insieme di questi fattori implica uno scostamento tra il moto previsto e quello effettivo. In altre parole, assegnato uno stato iniziale nominale \mathbf{X}_o^* , diverso dallo stato iniziale effettivo \mathbf{X}_o per via delle inevitabili incertezze, avranno luogo due traiettorie diverse: quella nominale, descritta da $\mathbf{X}^*(t)$, e quella effettiva, associata a $\mathbf{X}(t)$. Tale discrepanza rende necessari metodi di stima della traiettoria e di correzione degli errori, metodi alla cui base vi è il tracciamento del satellite attraverso stazioni a terra per $t > t_o$. Il tracciamento si basa sullo scambio di segnali tra le stazioni a terra e il satellite e le osservabili ricavate da questo processo corrispondono a quelle effettive, cioè associate allo stato effettivo del satellite, mentre lo stato nominale dà luogo alle osservabili attese. La differenza tra le osservabili effettive e quelle previste dal modello prende il nome di *residuo*. In formule:

$$\mathbf{Y}_i = G(\mathbf{X}_i, t_i) + \boldsymbol{\varepsilon}_i \quad i = 1, 2, \dots, l \quad (1.18)$$

essendo \mathbf{Y}_i il vettore delle osservabili effettive ottenute all'istante t_i , $G_i(\mathbf{X}_i, t_i) = \mathbf{Y}_i^*$ il vettore delle osservabili attese e $\boldsymbol{\varepsilon}_i$ il vettore degli errori nelle misure. Tutti e tre sono vettori $p \times 1$. La soluzione \mathbf{X}_i dell'Eq. (1.16) è esprimibile nella forma:

$$\mathbf{X}_i = \Theta(\mathbf{X}_o, t_o, t_i) \quad (1.19)$$

Sostituendo la (1.19) nella (1.18) si ottiene:

$$\mathbf{Y}_i = G[\Theta(\mathbf{X}_o, t_o, t_i), t_i] + \boldsymbol{\varepsilon}_i = \tilde{G}(\mathbf{X}_o, t_o, t_i) + \boldsymbol{\varepsilon}_i \quad (1.20)$$

Riscrivendo in una forma più compatta:

$$\mathbf{Y} = \tilde{G}(\mathbf{X}_o, t_o) + \boldsymbol{\varepsilon} \quad (1.21)$$

Osserviamo che la (1.21) è un'equazione vettoriale generalmente non lineare di $m = l \times p$ componenti scalari in n incognite di stato e m residui incogniti. Nell'ipotesi di $\boldsymbol{\varepsilon} = \mathbf{0}$, bastano n equazioni indipendenti per determinare le n incognite dello stato \mathbf{X} . Nel caso contrario di rumore casuale nelle misure, esistono diverse possibili soluzioni e occorre adottare un criterio per selezionare la soluzione migliore ai fini della stima orbitale.

Come spiegato precedentemente, il problema della stima dell'orbita di un satellite è non lineare, tuttavia, se la differenza tra la traiettoria prevista e quella effettiva è sufficientemente piccola durante l'intervallo temporale di interesse, è possibile effettuare una linearizzazione del problema. Definiamo rispettivamente i vettori dell'errore di stato e dell'errore di osservazione:

$$\mathbf{x}(t) = \mathbf{X}(t) - \mathbf{X}^*(t) \quad (1.22)$$

$$\mathbf{y}(t) = \mathbf{Y}(t) - \mathbf{Y}^*(t) \quad (1.23)$$

Sotto l'ipotesi di piccoli errori, possiamo espandere in serie di Taylor fino al primo ordine le Eq. (1.16) e (1.18) in un intorno dello stato nominale \mathbf{X}^* :

$$\dot{\mathbf{X}} = F(\mathbf{X}, t) = F(\mathbf{X}^*, t) + \left[\frac{\partial}{\partial \mathbf{X}} F(\mathbf{X}^*, t) \right] \mathbf{x} + o(\mathbf{x}) \quad (1.24)$$

$$\mathbf{Y}_i = G(\mathbf{X}_i, t_i) + \boldsymbol{\varepsilon} = G(\mathbf{X}_i^*, t_i) + \left[\frac{\partial}{\partial \mathbf{X}} G(\mathbf{X}^*, t) \right] \mathbf{x}_i + o(\mathbf{x}) + \boldsymbol{\varepsilon}_i \quad (1.25)$$

con $i = 1, 2, \dots, l$. Trascurando i termini di ordine superiore al primo, si giunge al seguente set di equazioni:

$$\dot{\mathbf{x}} = \left[\frac{\partial}{\partial \mathbf{X}} F(\mathbf{X}^*, t) \right] \mathbf{x} = A(t)\mathbf{x} \quad (1.26)$$

$$\mathbf{y}_i = \left[\frac{\partial}{\partial \mathbf{X}} G(\mathbf{X}^*, t) \right] \mathbf{x}_i + \boldsymbol{\varepsilon}_i = \tilde{H}_i \mathbf{x}_i + \boldsymbol{\varepsilon}_i \quad (1.27)$$

dove A e \tilde{H}_i sono matrici. La soluzione generale dell'Eq. (1.26) è la seguente:

$$\mathbf{x}(t) = \Phi(t, t_k)\mathbf{x}_k \quad (1.28)$$

essendo t_k un istante temporale di riferimento. L'operatore Φ è la matrice $n \times n$ di transizione di stato, le cui proprietà principali sono di seguito elencate:

1. $\dot{\Phi}(t, t_k) = A(t)\Phi(t, t_k)$
2. $\Phi(t_k, t_k) = I$, dove I è la matrice identità $n \times n$

Ponendo come istante di riferimento quello iniziale t_o e sostituendo la (1.28) nella (1.27), quest'ultima diventa:

$$\mathbf{y}_i = \tilde{H}_i \Phi(t_i, t_o)\mathbf{x}_o + \boldsymbol{\varepsilon}_i \quad (1.29)$$

o, in una forma più compatta:

$$\mathbf{y} = H\mathbf{x} + \boldsymbol{\varepsilon} \quad (1.30)$$

essendo \mathbf{y} e $\boldsymbol{\varepsilon}$ vettori $m \times 1$, \mathbf{x} un vettore $n \times 1$ e H una matrice $m \times n$, con $m = l \times p$. Poiché il sistema lineare ha più equazioni che incognite, esiste più di una soluzione e diversi metodi sono utilizzabili per selezionare quella migliore dal punto di vista della stima [3][1]. Uno strumento ampiamente utilizzato è il metodo dei minimi quadrati, formulato indipendentemente da Adrien-Marie Legendre (1752-1833) e Karl Friedrich Gauss (1777-1855), nonostante si attribuisca solitamente a quest'ultimo la paternità del metodo [4]. Come suggerisce il nome stesso, la soluzione migliore è quella che minimizza la norma euclidea $J = \|\boldsymbol{\varepsilon}\|$ del vettore degli errori di misura. Questo equivale a cercare il punto critico $\mathbf{X}_k^* = \hat{\mathbf{X}}_k$ di J , ossia quel punto per il quale si annulla il gradiente di J :

$$\left[\frac{\partial J}{\partial \mathbf{X}_k^*} \right]_{\hat{\mathbf{X}}_k} = \mathbf{0} \quad (1.31)$$

dove \mathbf{X}_k^* è un valore specificato dello stato in un'epoca arbitraria t_k . Perché $\hat{\mathbf{X}}_k$ sia punto di minimo, occorre verificare che in tale punto la matrice Hessiana di J sia definita positiva di segno [3]:

$$\delta \mathbf{X}_k^T \left[\frac{\partial^2 J}{\partial \mathbf{X}_k^* \partial \mathbf{X}_k^*} \right]_{\hat{\mathbf{X}}_k} \delta \mathbf{X}_k > 0 \quad (1.32)$$

A partire dai minimi quadrati, numerosi risultati sono stati ottenuti nell'ambito delle teorie della stima, ma in sostanza costituiscono aggiustamenti o estensioni dei fondamenti forniti da Gauss. Inoltre, è grazie alla maggiore precisione dei dati oggi disponibili se i metodi di stima sviluppati fino ad oggi sono in grado di esprimere le loro potenzialità [4].

Una volta ottenuta la migliore stima dello stato del satellite, questa è utilizzata per la predizione delle osservabili. I valori attesi sono poi confrontati con quelli effettivi forniti dagli strumenti di misura a terra e a bordo del satellite. Se i residui ottenuti da tale confronto esibiscono una distribuzione normale a media nulla, i modelli impiegati possono essere considerati accurati. Al contrario, se i modelli presentano inaccurately, i residui esibiscono una componente sistematica sovrapposta a quella casuale. I residui sono quindi utili per migliorare i suddetti modelli e di conseguenza le stime ottenute [3].

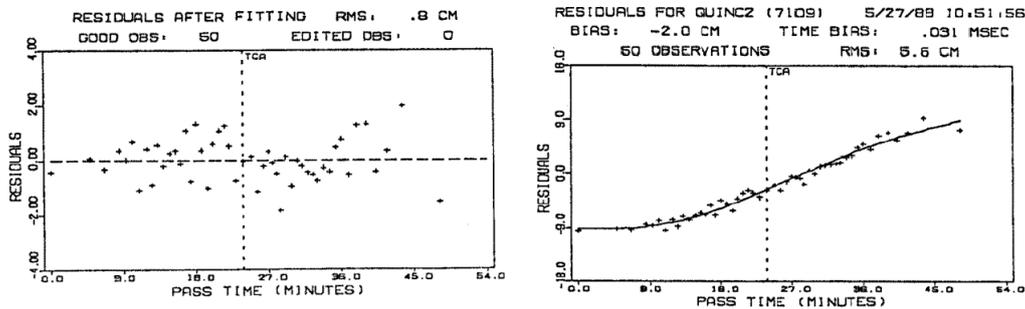


Figura 1.3: A sinistra, i residui esibiscono una distribuzione puramente casuale. I residui a destra, invece, presentano una componente sistematica che si aggiunge alle oscillazioni casuali [3]

Attraverso i modelli corretti, le nuove osservabili attese sono calcolate e confrontate con quelle effettive. Ne consegue una procedura iterativa di correzione dei modelli che si

arresta quando i residui soddisfano opportuni criteri, come quelli inerenti alla precisione ottenuta [5].

Possiamo concludere, dunque, che la determinazione d'orbita, secondo un approccio moderno, è un processo statistico e iterativo il quale consiste nella ricerca della migliore stima dello stato di un satellite in un dato istante t a partire da misurazioni e modelli affetti da errori e incertezze [1][4].

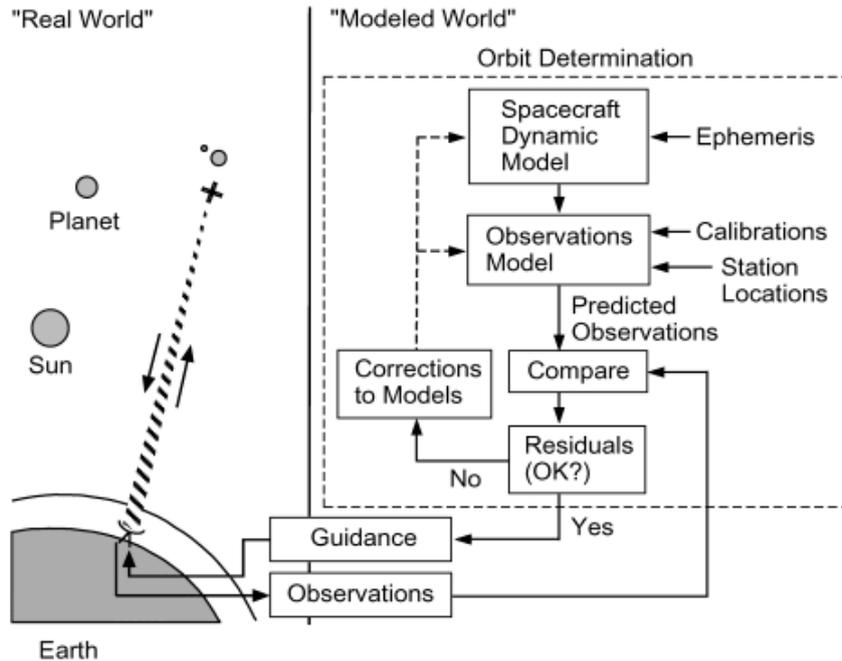


Figura 1.4: Schema della procedura iterativa di stima del moto orbitale di un satellite [5]

Tali errori ed incertezze seguono processi casuali descrivibili per mezzo di distribuzioni statistiche, distribuzioni che possono essere parametriche, cioè descritte da parametri. Sia x una variabile casuale, con $\xi < x < \xi + d\xi$, allora la probabilità dell'evento associato a x è:

$$P(\xi) = p(\xi)d\xi \quad (1.33)$$

essendo p la funzione densità di probabilità.

Secondo la definizione assiomatica di probabilità, formulata da Andrej Nikolaevic Kolmogorov (1903-1987), l'intera distribuzione di probabilità $p(x)$ deve rispettare la seguente condizione:

$$\int_{-\infty}^{+\infty} p(\xi)d\xi = 1 \quad (1.34)$$

La media \bar{x} è definita come il valore atteso di un evento su tutto l'intervallo dei possibili esiti o, in alternativa, come l'integrale (o la somma) di tutti i valori assunti dalla variabile casuale x , ciascuno moltiplicato per la sua probabilità.

$$\bar{x} = E(x) = \int_{-\infty}^{+\infty} \xi p(\xi)d\xi \quad (1.35)$$

Poiché calcolata sulla totalità dei possibili esiti, essa è detta anche media della popolazione. Esiste poi la media campionaria, basata su un sottoinsieme dei possibili esiti:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N \xi_i \quad (1.36)$$

La legge dei grandi numeri afferma che, all'aumentare del numero di campioni, la media campionaria approssima sempre meglio quella della popolazione.

C'è poi la varianza, la quale indica di quanto si scosta in media la variabile casuale x dal suo valore medio:

$$\sigma^2 = E\{[x - \bar{x}]^2\} = \int_{-\infty}^{+\infty} (\xi - \bar{x})^2 p(\xi) d\xi \simeq \frac{1}{N} \sum_{i=1}^N (\xi_i - \bar{x})^2 \quad (1.37)$$

Si definisce deviazione standard σ la radice quadrata della varianza [4]:

$$\sigma = \sqrt{\sigma^2} \quad (1.38)$$

1.1.2 Tracking e acquisizione dati

Nella precedente sezione si è accennato al ruolo fondamentale dell'operazione di tracking nel processo di stima del moto orbitale di un satellite.

Il tracking è realizzato attraverso una varietà di tecniche radiometriche oppure ottiche. Per esempio, durante la fase di crociera di un trasferimento interplanetario si può fare uso di tecniche radiometriche. Queste sono utilizzate anche per fornire informazioni accurate sull'orbita, qualora siano necessarie correzioni della traiettoria durante lo svolgimento della missione, e potrebbero essere utilizzate insieme alle immagini ottiche nella fase di approccio per diminuire l'incertezza della posizione relativa all'obiettivo. Le tecniche radiometriche si basano sullo scambio di segnali tra una stazione a terra GS (*Ground Station*) e il satellite S/C (*Spacecraft*). Tali segnali sono contenuti in bande di frequenza pre-allocate a livello internazionale.

Band	Uplink Frequency (MHz)	Downlink Frequency (MHz)
S	2110–2120	2290–2300
X	7145–7190	8400–8450
Ka	34,200–34,700	31,800–32,300

Figura 1.5: Principali bande di frequenza impiegate per i segnali radio [5]

Si tende a preferire frequenze maggiori per via delle migliori performance di comunicazione e per aumentare l'accuratezza delle misure radiometriche grazie alle minori lunghezze d'onda, oltre a ridurre gli effetti dell'interazione con le particelle cariche della ionosfera e del plasma solare e interplanetario [5].

Si definisce uplink il segnale inviato dalla GS allo S/C; viceversa, il downlink è il segnale inviato dallo S/C alla GS. Esistono diverse modalità di tracking, di seguito illustrate.

La prima è quella cosiddetta *one-way*, nella quale il satellite trasmette il downlink a terra. La frequenza del downlink è generalmente diversa da quella di trasmissione della GS per evitare interferenza. Il limite principale del one-way tracking è la stabilità in frequenza dell'oscillatore a bordo che produce il segnale, solitamente inferiore a quella della strumentazione a terra, motivo per cui è preferibile la modalità *two-way*. Nel two-way tracking, la GS invia l'uplink, questo è ricevuto dal transponder a bordo dello S/C e per mezzo di un Phase-Locked Loop, o PLL, aggancia la fase della portante e produce un downlink coerente in fase all'uplink. La frequenza del downlink è inoltre amplificata di un fattore G , o *transponding ratio*, necessario per la distinzione dall'uplink. Il downlink è infine ricevuto dalla stessa stazione che ha trasmesso l'uplink, di conseguenza la modalità two-way è limitata dalla finestra temporale di osservabilità dello S/C. Tale limitazione è superata per mezzo del *three-way* tracking, il quale differisce dal two-way per il fatto che il downlink è ricevuto da una seconda GS, la cui frequenza standard potrebbe differire da quella della prima stazione.

Una volta ricevuto, il downlink subisce un trattamento al fine di estrarne le informazioni correlate e utilizzarle per risalire alle osservabili. Di seguito è descritta la procedura generale di acquisizione del downlink operata dal Deep Space Network della NASA e schematizzata in Fig. (1.6).

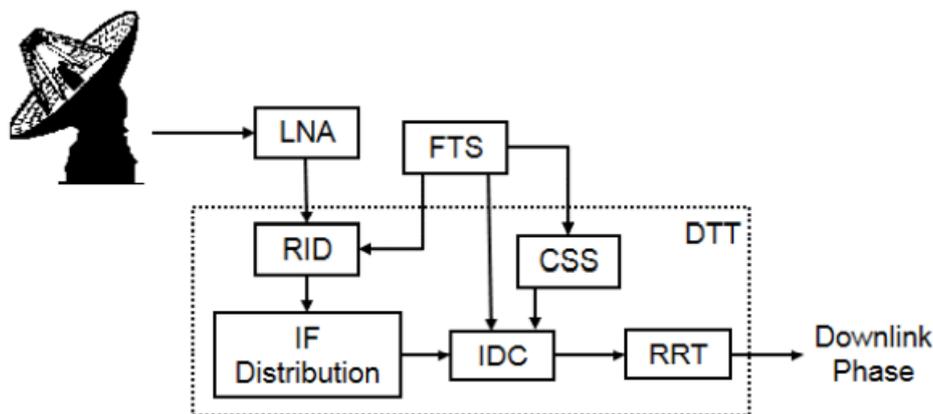


Figura 1.6: Sequenza di acquisizione del downlink da parte di una stazione del Deep Space Network [6]

Dopo essere stata captata dall'antenna, la portante del downlink passa per il Low Noise Amplifier LNA e accede al Downlink Tracking and Telemetry Subsystem DTT, il quale risiede in parte nell'antenna e in parte nel Signal Processing Center SPC. La prima tappa del DTT è il Radio-frequency to Intermediate-frequency Downconverter RID, situato nell'antenna, dove ha luogo il battimento della portante del downlink, cioè la diminuzione della sua frequenza per mezzo della frequenza di riferimento dell'uplink fornita dal Frequency and Timing Subsystem FTS. La portante battuta passa dunque al SPC, iniziando dall' Intermediate-frequency to Digital Converter IDC, dove la frequenza del segnale è portata al valore finale di circa 200 MHz attraverso una combinazione di conversioni di tipo up e down. In particolare, lo stadio finale di downconversion utilizza la frequenza fornita dal Channel-Select Synthesizer CSS, frequenza impostata all'inizio di ogni acquisizione sulla base della banda di frequenze attesa e che rimane costante durante l'acquisizione stessa. Tutte le frequenze utilizzate per le conversioni sono fornite dal FTS e caratterizzate da alta stabilità. In seguito, l'IDC converte il segnale battuto dal formato analogico a quello digitale.

La procedura fin qui descritta non presenta rami in retroazione, motivo per cui i dati così acquisiti sono definiti *open loop*. Il vantaggio di questi segnali è la possibilità di registrare i dati nella memoria di un computer per elaborarli a posteriori e più di una volta, così da migliorare la ricostruzione del segnale filtrato dal rumore e la conseguente estrazione di informazioni da esso [6].

1.1.3 Osservabili

Una volta esplorate le modalità di tracciamento del satellite e l'acquisizione delle informazioni associate ai segnali scambiati con le stazioni a terra, procediamo con un'analisi delle osservabili estraibili da tali informazioni e che saranno utilizzate per risalire allo stato del satellite. Le osservabili consistono solitamente in misure di range, range-rate, ascensione retta e declinazione. Il range ρ è definito come la distanza tra la stazione a terra e il satellite:

$$\rho = \|\boldsymbol{\rho}\| = \|\mathbf{r} - \mathbf{r}_{st}\| \quad (1.39)$$

dove \mathbf{r} è la posizione del satellite e \mathbf{r}_{st} quella della stazione a terra, entrambe definite in un opportuno sistema di riferimento. Il modello di predizione del range dipende dalla modalità di tracciamento.

One-way range. Il satellite trasmette un segnale all'istante t_T e la stazione a terra lo riceve all'istante t_R . Poiché il segnale viaggia di moto rettilineo uniforme con velocità pari a quella delle onde elettromagnetiche nel vuoto c , il range è calcolabile come:

$$\tilde{\rho} = c(t_R - t_T) \quad (1.40)$$

La notazione $\tilde{\rho}$ indica che il modello non fornisce un vero range, bensì uno *pseudo-range*. Il motivo è che gli orologi a bordo del satellite e a terra potrebbero non essere sincronizzati, oltre a presentare problemi di instabilità nella frequenza di oscillazione. Queste considerazioni portano al seguente modello di misura del tempo t di un orologio:

$$t = T + a + b(T - T_o) + \varepsilon' \quad (1.41)$$

essendo T il tempo vero, a un termine costante di scostamento dal tempo vero, b il coefficiente del contributo lineare di scostamento, T_o un tempo di riferimento e ε' il contributo di termini non lineari e stocastici. Calcolando l'Eq. (1.41) per gli istanti t_R e t_T e sostituendo nell'Eq. (1.40):

$$\begin{aligned} \tilde{\rho} &= c(T_R - T_T) + c(a_R - a_T) + c(b_R - b_T)(T - T_o) + \varepsilon'_R - \varepsilon'_T \Rightarrow \\ \Rightarrow \tilde{\rho} &= \rho + c(a_R - a_T) + c(b_R - b_T)(T - T_o) + \varepsilon \end{aligned}$$

dove ρ è il range vero. Un modello più accurato tiene anche conto dell'errore di range dovuto al delay $\delta\rho_{atm}$ che ha luogo quando il segnale attraversa l'atmosfera:

$$\tilde{\rho} = \rho + c(a_R - a_T) + c(b_R - b_T)(T - T_o) + \varepsilon + \delta\rho_{atm} \quad (1.42)$$

Dalla differenza tra lo pseudo-range osservato e quello calcolato per mezzo dell'Eq. (1.42) si ottiene un primo residuo per la determinazione d'orbita [1].

Two-way range. L'uplink consiste in una portante sulla quale è modulato un codice, detto range code. Una volta che l'uplink è tracciato dal transponder a bordo del satellite, un PLL aggancia la fase della sua portante per produrre un segnale coerente. Quest'ultimo è utilizzato per demodulare l'uplink al fine di estrarre il range code e filtrarlo. Successivamente il range code filtrato è modulato sulla portante del downlink per la trasmissione a terra. Qui il PLL della stazione ricevente aggancia la fase della portante e produce un segnale coerente ad essa per la demodulazione. Il range code così estratto è confrontato con quello trasmesso per determinare il tempo di percorrenza [5]. Poiché il segnale percorre due volte la distanza che separa il satellite dalla stazione a terra, si parla di *round-trip range*, cioè range di andata e ritorno:

$$\rho_{rt} = c(t_R - t_T) \quad (1.43)$$

Adottando l'Eq. (1.41), si giunge alla seguente espressione:

$$\rho_{rt} = c(T_R - T_T) + b(T_R - T_T) + \varepsilon + \delta\rho_{atm} \quad (1.44)$$

Osserviamo che, essendo la stazione trasmittente coincidente con quella ricevente, $a_R = a_T$, dunque il contributo di scostamento costante si annulla con un conseguente vantaggio per quanto riguarda la bontà delle predizioni. Non avendo a disposizione l'informazione sull'istante di arrivo del segnale al satellite, il calcolo del ρ_{tr} atteso non è una procedura diretta, ma iterativa. Più nello specifico, la procedura è suddivisa in due iterazioni:

1. stima del tempo di arrivo al satellite e dell'uplink range
2. stima del tempo di arrivo a terra e del downlink range

Quelle finora descritte sono procedure generali, ma sotto opportune condizioni è possibile operare delle semplificazioni [1].

Il range-rate è definito come la proiezione della velocità del satellite lungo la linea di vista che congiunge il satellite e la stazione di tracking o, in altre parole, la velocità radiale del satellite relativa alla stazione:

$$\dot{\rho} = \frac{1}{\rho} \dot{\mathbf{r}} \cdot \boldsymbol{\rho} \quad (1.45)$$

Un tipico modello di range-rate è il seguente:

$$\dot{\rho}(t) = \dot{r}(t) + \omega_e r_s \cos(\delta) \sin(\omega_e t + \phi + \lambda_s - \alpha) \quad (1.46)$$

dove ω_e è la velocità media di rotazione della Terra, δ la declinazione, α l'ascensione retta, ϕ un contributo di fase funzione dell'istante di tempo considerato e λ_s la longitudine della stazione. L'Eq. (1.46) mostra come il range-rate sia la sovrapposizione di due contributi [5]:

1. la velocità geocentrica del satellite \dot{r} , cioè relativa al centro della Terra

2. l'oscillazione dovuta al moto di rotazione terrestre al quale partecipa la stazione e la cui ampiezza e fase sono funzione, rispettivamente, della declinazione e dell'ascensione retta

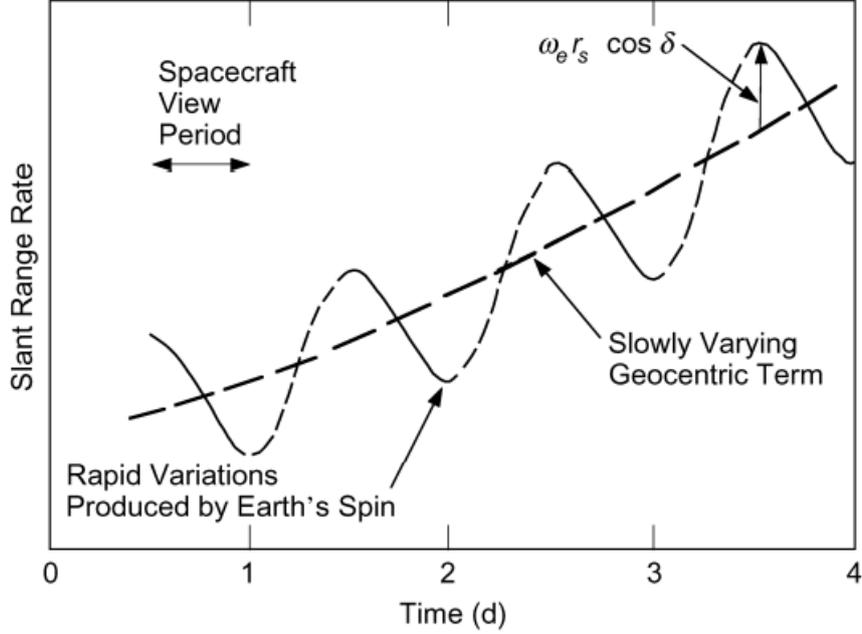


Figura 1.7: Grafico del range-rate. Notare la sovrapposizione del termine sinusoidale e di quello geocentrico [5]

Le tecnologie attuali misurano il range-rate attraverso due metodi: trasmissione di impulsi o di segnali a frequenza nota. La seguente trattazione è riferita al one-way tracking.

Repeated Pulse Transmission. Il satellite trasmette impulsi di durata δt_T agli istanti t_{T1}, t_{T2}, \dots e l' i -esimo istante di ricezione è pari a:

$$t_{Ri} = t_{Ti} + \frac{\rho_i}{c} \quad (1.47)$$

Dati due impulsi consecutivi, l'intervallo di tempo tra la ricezione di uno e dell'altro è:

$$t_{R2} - t_{R1} = \delta t = \delta t_T + \frac{\rho_2 - \rho_1}{c} = \delta t_T + \frac{\delta \rho}{c} = \delta t_T \left(1 + \frac{1}{c} \frac{\delta \rho}{\delta t_T} \right) \quad (1.48)$$

Transmitter Beacon. Il satellite trasmette un downlink a frequenza nota f_T e ricevuto a terra con frequenza f_R . Qui il downlink è mixato con un segnale di riferimento a frequenza f_G , solitamente pari a f_T , ottenendo un segnale con due componenti di frequenza $f_G + f_R$ e $f_G - f_R$. Successivamente un filtro passa-basso restituisce la sola componente a frequenza $f_G - f_R$, dopodiché il ricevitore ne conta il numero $N_{1,2}$ di cicli tra gli istanti 1 e 2:

$$N_{1,2} = \phi \pmod{2\pi} = \int_{t_{R1}}^{t_{R2}} (f_G - f_R) dt \quad (1.49)$$

essendo $\phi \pmod{2\pi}$ la fase in modulo 2π del segnale. Ponendo $t_{Ri} = t_{Ti} + \Delta t_i = t_{Ti} + \rho_i(t_{Ri})/c$ e f_G costante, l'Eq. (1.49) diventa:

$$N_{1,2} = f_G \left(t_{T2} - t_{T1} + \frac{\rho_2 - \rho_1}{c} \right) - \int_{t_{R1}}^{t_{R2}} f_R dt \quad (1.50)$$

Per la conservazione della fase e assumendo la frequenza trasmessa f_T costante,

$$\int_{t_{R1}}^{t_{R2}} f_R dt = \int_{t_{T1}}^{t_{T2}} f_T dt = f_T(t_{T2} - t_{T1}) \quad (1.51)$$

quindi,

$$N_{1,2} = f_G \left(t_{T2} - t_{T1} + \frac{\rho_2 - \rho_1}{c} \right) - f_T(t_{T2} - t_{T1}) \quad (1.52)$$

Se $f_G = f_T$,

$$N_{1,2} = f_T \frac{\rho_2 - \rho_1}{c} = f_T \frac{\delta\rho}{c} \Rightarrow \frac{N_{1,2}}{\delta t} = \frac{f_T}{c} \frac{\delta\rho}{\delta t} \quad (1.53)$$

Differenziando rispetto al tempo l'Eq. (1.49) e sfruttando l'Eq. (1.53),

$$f_R = f_T \left(1 - \frac{1}{c} \frac{\delta\rho}{\delta t} \right) \quad (1.54)$$

Notare la simmetria delle Eq. (1.48) e (1.54) dovuta al termine

$$\frac{1}{c} \frac{\delta\rho}{\delta t} \quad (1.55)$$

il quale approssima tanto meglio il range-rate $\dot{\rho}$ istantaneo quanto più piccolo è l'intervallo di tempo δt . Esso è il contributo dell'effetto Doppler, cioè la variazione della frequenza del segnale a causa del moto relativo tra stazione e satellite. Nel caso di transmitter beacon, dall'Eq. (1.54), la variazione di frequenza per effetto Doppler, o frequenza Doppler, è pari a:

$$f_D = \frac{f_T}{c} \frac{\delta\rho}{\delta t} \quad (1.56)$$

Per misurazioni in modalità two-way o three-way, occorre tenere conto del Doppler shift di frequenza dell'uplink e del downlink, oltre allo scostamento in frequenza tra i due segnali attraverso un transponding ratio G per la loro distinzione:

$$f_D = \frac{2}{G} \frac{f_T}{c} \frac{\delta\rho}{\delta t} \quad (1.57)$$

Manipolando l'Eq. (1.52) otteniamo un'ulteriore interpretazione del numero di cicli $N_{1,2}$:

$$\rho_2 = \rho_1 + N_{1,2} \frac{c}{f_G} - \frac{c}{f_G} (f_G - f_T)(t_{T2} - t_{T1}) \quad (1.58)$$

L'Eq. (1.58) mostra come il numero di cicli, oltre a fornire informazioni sul range-rate, possa essere utilizzato anche per ricostruire la variazione di range tra t_1 e t_2 [1].

Un singolo passaggio del satellite nel campo di vista della stazione a terra consente di ottenere informazioni sul range-rate, la declinazione e l'ascensione retta. Al contrario, la misurazione delle componenti di velocità normali alla linea di vista richiedono più passaggi [5].

La bontà delle osservabili è funzione di molte variabili e frutto di un compromesso tra i vantaggi e gli svantaggi dei fattori che concorrono all'ottenimento e all'elaborazione delle osservabili stesse. Un primo aspetto da tenere in considerazione è la disponibilità dei dati da parte dei sensori che li forniscono. C'è poi la quantità: un numero esiguo di osservabili complica la determinazione d'orbita e può portare alla formulazione di modelli scorretti. Inoltre, una maggiore densità di dati, cioè la stessa quantità di dati ottenuta da un numero minore di stazioni a terra, incrementa l'accuratezza della stima orbitale. Ulteriori fattori sono la posizione relativa tra satellite e stazione, la tipologia dei dati e l'osservabilità del satellite [4].

1.1.4 Fonti di errore

Abbiamo visto come gli errori giochino un ruolo importante nel processo di determinazione d'orbita. Le fonti da cui essi scaturiscono non sono da ricercare solo nei fenomeni fisici, come il rumore nelle misurazioni, ma anche nella formulazione dei modelli matematici del mondo reale e nell'approssimazione dei valori numerici [3]. L'errore nelle misurazioni può essere definito in prima battuta come lo scostamento delle osservabili attese da quelle effettive e consiste in tre contributi:

1. bias: scostamento costante dal valore reale
2. rumore: variazioni casuali intorno al valore medio
3. drift: lenta e imprevedibile variazione del valore medio in un dato intervallo di tempo

Nel caso degli errori nei modelli matematici, esiste a sua volta un modello degli errori nella formulazione della dinamica del sistema, noto come *rumore di processo*, il quale quantifica l'abilità nel modellare accuratamente la dinamica prima e dopo un certo istante di tempo [4]. Non bisogna dimenticare poi gli errori nelle procedure computazionali di troncamento e arrotondamento dei valori numerici e di integrazione numerica delle equazioni [1]. Molte fonti di errore lasciano tracce distinte, mentre altre sono molto simili tra loro rendendo difficile, se non impossibile, il riconoscimento al fine di effettuarne la separazione dai dati originali [3].

Una prima fonte di errori risiede nell'atmosfera. I segnali impiegati occupano diverse bande di frequenza e la frequenza stabilisce il tipo di interazione con l'atmosfera, in particolare con gli strati della troposfera e della ionosfera. La troposfera è lo strato dell'atmosfera terrestre a diretto contatto con la superficie e si estende fino ad una quota di circa 10-11 km. Qui hanno luogo la vita e gli eventi meteorologici. La troposfera provoca rifrazione con associato rallentamento del segnale, detto anche *propagation delay*. Il rallentamento si traduce in un aumento fittizio di range:

$$\delta\rho = 10^{-6} \int N ds \quad (1.59)$$

essendo s una variabile di integrazione lungo il percorso e N la *refrattività*. Quest'ultima è funzione dell'indice di rifrazione n del mezzo di propagazione:

$$N = 10^6(n - 1) \quad (1.60)$$

In particolare, la refrattività è la somma di due contributi:

$$N = N_d + N_w \quad (1.61)$$

La refrattività N_d è riferita ad una troposfera priva di umidità (il pedice d è l'iniziale di *dry*, cioè asciutto) e costituisce circa il 90% della refrattività totale. Essa assume che la troposfera si comporti come un gas ideale in equilibrio idrostatico ed è proporzionale alla pressione in corrispondenza della superficie terrestre. La componente N_w tiene conto dell'umidità nella troposfera (il pedice w sta per *wet*, cioè umido) ed è più difficile da modellare a causa della forte variabilità nello spazio e nel tempo della densità di vapore acqueo.

Minore è l'elevazione della linea di vista, maggiore sarà il delay troposferico a causa del maggior spessore di atmosfera da attraversare [1][5]. A differenza della banda ottica, la troposfera è un mezzo non dispersivo per le onde radio in quanto elettricamente neutro.

La ionosfera è la porzione di atmosfera caratterizzata dalla presenza di ioni ed elettroni liberi e si colloca a quote superiori a 80 km. Essa consiste di vari strati, ma non tutti sono completamente distinti. La ionosfera non ha effetti significativi sui segnali ottici, mentre costituisce un mezzo dispersivo alle frequenze radio. I segnali che la attraversano subiscono un time delay quantificabile come:

$$\delta t = \frac{\alpha}{f^2} \quad (1.62)$$

dove f è la frequenza e α funzione del Total Electron Content TEC, a sua volta funzione della posizione e del tempo. Quando si fa uso di una sola frequenza, modelli predittivi della ionosfera possono essere utilizzati per correggere le misure di range, tuttavia una maggiore accuratezza è data dalla stessa misura effettuata con segnali a frequenze diverse. Ad esempio, in modalità *dual frequency*, date due componenti di frequenza f_1 e f_2 , il valore finale di range ρ_c è dato dalla combinazione lineare dei range misurati attraverso le singole frequenze:

$$\rho_c = \gamma_1 \rho_1 + \gamma_2 \rho_2 \quad (1.63)$$

essendo

$$\gamma_1 = \frac{f_1^2}{f_1^2 - f_2^2} \quad (1.64)$$

$$\gamma_2 = \frac{f_2^2}{f_1^2 - f_2^2} \quad (1.65)$$

La combinazione lineare rimuove la maggior parte, se non la totalità, degli effetti della variabilità del TEC ionosferico, motivo per cui ρ_c è anche detto *ionosphere-free range*, ma al prezzo di un maggiore rumore nelle misure [1]. Il delay ionosferico, inoltre, è maggiore di giorno che di notte di circa un ordine di grandezza e aumenta al diminuire dell'elevazione della linea di vista, sempre a causa del maggior spessore di atmosfera che il segnale deve

attraversare. I delay atmosferici, in generale, dipendono dal ciclo dell'attività solare, essendo la densità atmosferica proporzionale ad essa [5].

Oltre a quelli atmosferici, ulteriori delay sono quelli causati dai fenomeni relativistici [1].

Esiste poi l'instabilità in frequenza degli orologi e degli oscillatori che producono i segnali alle frequenze di riferimento. Un errore nella frequenza di riferimento per il battimento del downlink, ad esempio, produce una varianza nella stima del range-rate pari a:

$$\Delta\dot{\rho} = c \frac{\Delta f}{f} \quad (1.66)$$

Solitamente, la modalità two-way di tracking è più precisa di quella one-way poiché la frequenza di riferimento della stazione a terra è più stabile di quella a bordo del satellite. La misura della stabilità della frequenza di riferimento, o frequenza standard, nel dominio del tempo è la varianza di Allan.

Quanto alla strumentazione, gli errori hanno origine da fattori quali la fisica degli strumenti impiegati, i compromessi nel loro progetto e la richiesta di più funzionalità in un unico dispositivo.

Altri errori sono quelli relativi alla definizione della posizione delle stazioni di tracking nel sistema di riferimento adottato. Un errore nella longitudine λ_s della stazione produce lo stesso errore nell'ascensione retta, mentre un errore nella distanza r_s della stazione rispetto al centro della Terra, porta ad un errore nella declinazione:

$$\Delta\delta = \frac{\Delta r_s}{r_s} \cot \delta \quad (1.67)$$

Occorre poi tenere in considerazione le variazioni nell'orientamento della Terra, così come nei suoi moti di rotazione e di rivoluzione, a causa dell'influenza gravitazionale del Sole e della Luna, della non-sfericità della Terra stessa e degli eventi che ne modificano le caratteristiche superficiali o negli strati più interni.

Ultima, ma non meno importante, fonte di errori è il rumore termico. Il rumore termico è l'insieme dei segnali generati dal naturale movimento di agitazione termica degli elettroni (è noto che cariche elettriche in accelerazione emettono onde elettromagnetiche) e il suo effetto cresce con l'aumento della temperatura assoluta. Il rumore termico può essere modellato matematicamente come un rumore gaussiano bianco o AWGN, dall'inglese Additive White Gaussian Noise. L'aggettivo *gaussiano* indica che l'ampiezza del rumore è esprimibile attraverso una distribuzione di probabilità gaussiana, anche detta distribuzione normale. Una distribuzione normale di probabilità è definita da due parametri: la media μ e la deviazione standard σ . Nel caso del rumore termico, $\mu = 0$.

L'aggettivo *bianco* indica l'assenza di periodicità nel tempo, cioè i valori assunti sono completamente casuali e non esiste alcuna correlazione tra di essi, e che la densità spettrale di potenza è costante per tutte le frequenze. L'origine dell'aggettivo *bianco* risiede nel fatto che una radiazione elettromagnetica avente le suddette caratteristiche apparirebbe di colore bianco nella banda di frequenze della luce visibile [7].

Infine, l'aggettivo *additivo* indica che il rumore termico è algebricamente sommato al segnale originale:

$$s'(t) = s(t) + n(t) \quad (1.68)$$

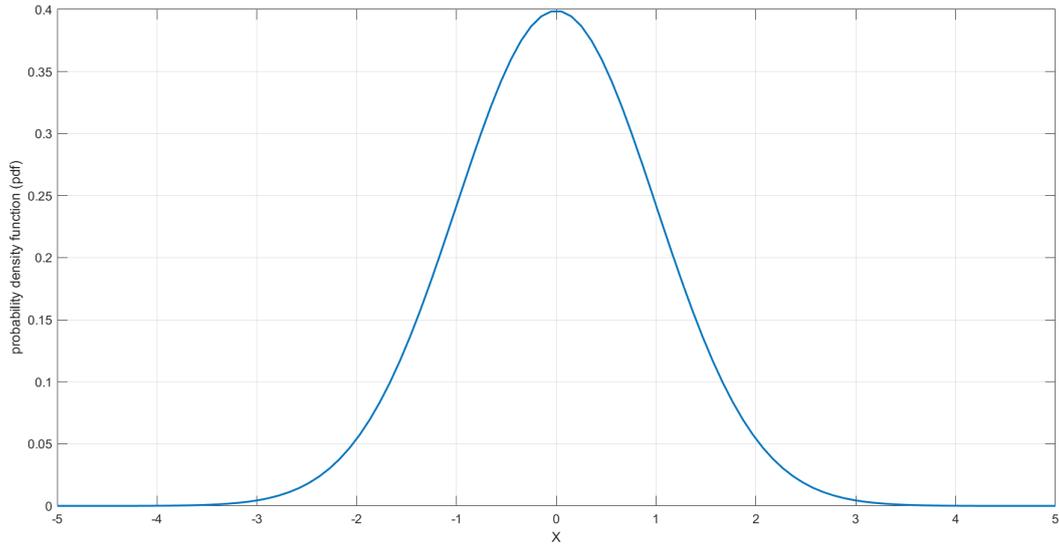


Figura 1.8: Distribuzione normale di probabilità con $\mu = 0$ e $\sigma = 1$. L'area sottesa è pari a 1



Figura 1.9: Spettro del rumore bianco. Fonte: https://it.wikipedia.org/wiki/Rumore_bianco

essendo $s(t)$ il segnale originale e $n(t)$ il rumore termico.

Avendo introdotto il rumore termico, può essere utile definire il *carrier-to-noise ratio* CNR :

$$CNR = \frac{P_C}{P_N} \quad (1.69)$$

P_C è la potenza della portante, o *carrier*, del segnale:

$$P_C = A^2 \quad (1.70)$$

P_N è la potenza associata al rumore termico:

$$P_N = 2\sigma^2 = N_o f_s \quad (1.71)$$

dove N_o è la densità spettrale di potenza, costante per il rumore bianco, e f_s è la frequenza di campionamento del segnale. Talvolta il CNR è espresso in dB:

$$CNR_{dB} = 10 \log_{10} (CNR) \quad (1.72)$$

Sostituendo le Eq. (1.70) e (1.71) nella (1.72) si giunge alla seguente formula:

$$CNR_{dB} = 10 \log_{10} \left(\frac{A^2}{N_o f_s} \right) = 10 \log_{10} \left(\frac{A^2}{N_o} \right) - 10 \log_{10} f_s = CN0 - 10 \log_{10} f_s$$

essendo $CN0$ il *carrier-to-noise density ratio*, espresso in dB-Hz.

In alcuni casi, le misure ottenute dal tracking possono essere calibrate per mezzo dei dati inviati dalla costellazione di satelliti GPS, acronimo di *Global Positioning System* [5].

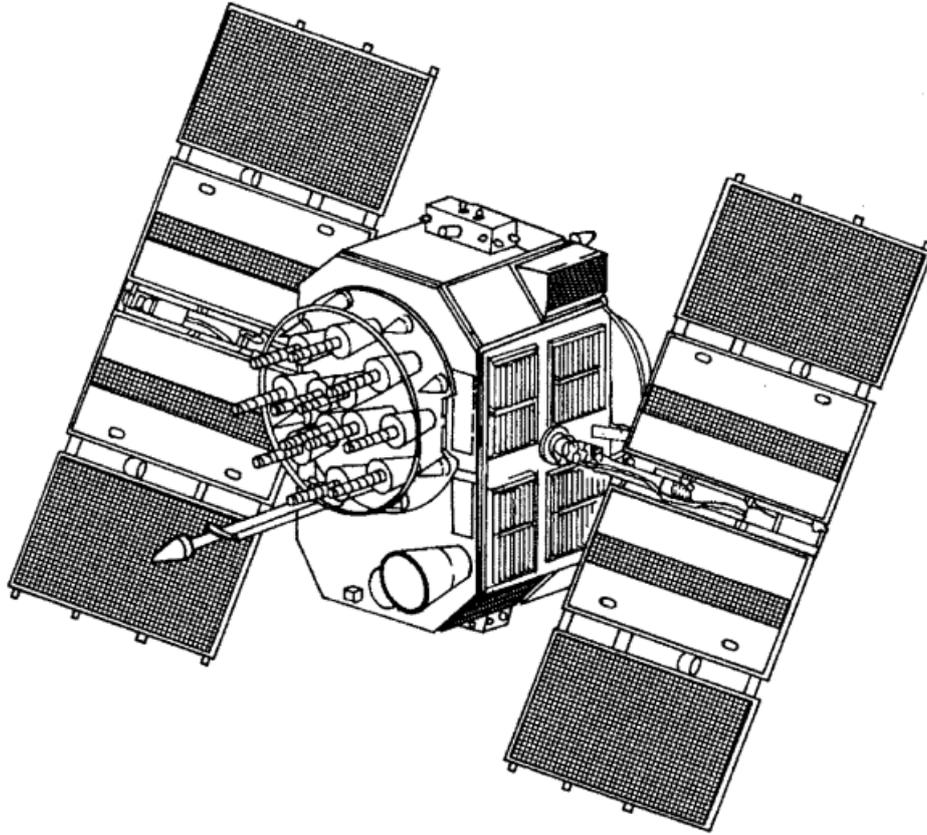


Figura 1.10: Satellite GPS del blocco II. La costellazione GPS consiste in più di 24 satelliti, distribuiti su 6 orbite aventi stessa quota e inclinazione, ma diversa ascensione retta. Il GPS è proprietà del Dipartimento di Difesa degli Stati Uniti d'America e in passato il segnale era soggetto alla Selected Availability SA, cioè una degradazione intenzionale per gli utenti civili. Nel 2000 la SA è stata rimossa, offrendo anche alle utenze non militari le massime potenzialità del sistema GPS [1]

1.2 Digital Signal Processing

1.2.1 Rappresentazione dei segnali

Nel processo di determinazione d'orbita è stato spiegato come le osservabili siano ottenute dallo scambio di segnali tra una stazione a terra e il satellite. In questa sezione si vuole esaminare più nel dettaglio cosa sia un segnale e fornire alcune nozioni di base della teoria dei segnali. Un segnale è una funzione del tempo $s(t)$ che descrive la variazione temporale di una qualunque grandezza fisica utilizzata per rappresentare o trasmettere informazioni. Nel caso delle telecomunicazioni tra satellite e terra, la grandezza fisica in esame è il campo elettrico di un'onda elettromagnetica. Considereremo quindi il seguente segnale sinusoidale:

$$E(t) = A_E(t) \cos \theta_E(t) \quad (1.73)$$

dove E è l'intensità del campo elettrico in un punto dello spazio, $A_E(t)$ è l'ampiezza dell'oscillazione all'istante t , e $\theta(t)$ è la fase del segnale, definita nel modo seguente:

$$\theta(t) = \int_{t_0}^t 2\pi f(\tau) d\tau \quad (1.74)$$

La funzione integranda f è la frequenza del segnale, misurata in Hz. Dalla precedente definizione di fase, segue che:

$$f(t) = \frac{1}{2\pi} \frac{d}{dt} \theta(t) \quad (1.75)$$

In altre parole, la frequenza esprime la rapidità con cui varia la fase del segnale nel tempo.

Solitamente, quando un'onda elettromagnetica è ricevuta da un'antenna, il campo elettrico dà luogo ad una tensione nel circuito a valle dell'antenna stessa, motivo per cui possiamo considerare da questo punto in poi il segnale sinusoidale di tensione:

$$V(t) = A_V(t) \cos \theta_V(t) \quad (1.76)$$

dove $A_V \neq A_E$ a causa del guadagno dell'antenna e la fase $\theta_V(t)$ è generalmente diversa da quella trasmessa per via del rumore, delle caratteristiche del canale di comunicazione e di altri fattori.

Quello appena illustrato è un esempio di rappresentazione temporale di un segnale, cioè nel dominio del tempo. La rappresentazione temporale rivela informazioni sull'inizio e la fine del segnale, la sua intensità ed evoluzione nel tempo e come la sua energia è distribuita nel tempo.

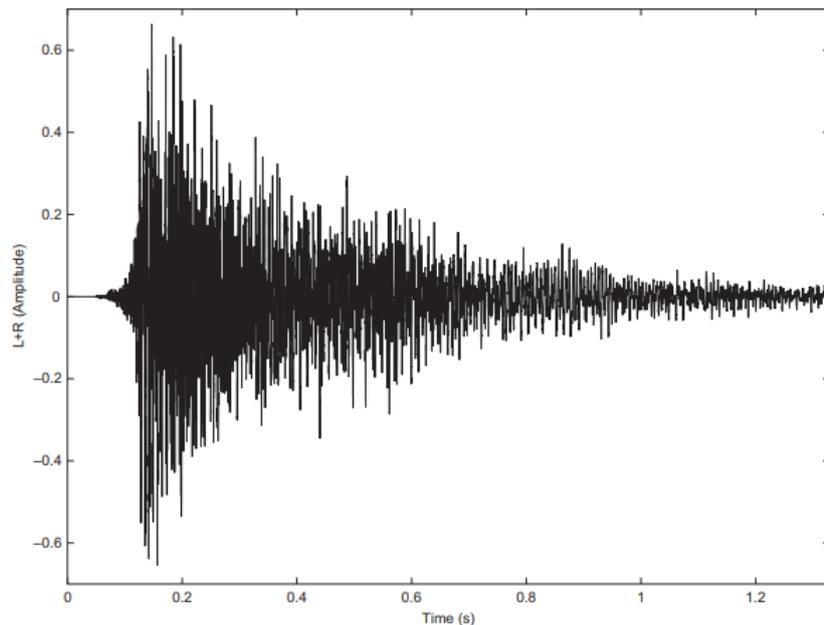


Figura 1.11: Rappresentazione di un segnale nel dominio del tempo [8]

La rappresentazione nel tempo tuttavia non è completa, cioè non fornisce tutte le informazioni di un segnale, perciò sono necessarie ulteriori rappresentazioni. Si deve a Jean Baptiste Joseph Fourier (1768-1830) la formulazione di un teorema secondo cui un segnale di intensità finita che soddisfa specifiche condizioni equivale alla combinazione lineare di seni e coseni, o esponenziali complessi, a differenti frequenze. Per ogni frequenza, il coefficiente che moltiplica la rispettiva senoide è una quantità complessa data dalla funzione della frequenza $S(f)$, detta spettro del segnale e ottenuta per mezzo della trasformata integrale di Fourier:

$$S(f) = \mathcal{F}\{s(t)\} = \int_{-\infty}^{+\infty} s(t)e^{-j2\pi ft} dt \quad (1.77)$$

In altre parole, la trasformata di Fourier consente il passaggio dal dominio del tempo a quello delle frequenze e lo spettro è la rappresentazione del segnale in quest'ultimo dominio. Nonostante l'assenza di informazioni temporali, lo spettro mostra quali frequenze sono contenute nel segnale, le loro rispettive intensità, la minima e la massima frequenza e di conseguenza la loro differenza, cioè l'ampiezza di banda del segnale. Poiché, in generale, lo spettro di un segnale è una quantità complessa, a partire dal suo modulo e argomento si definiscono, rispettivamente, lo spettro di ampiezza e lo spettro di fase.

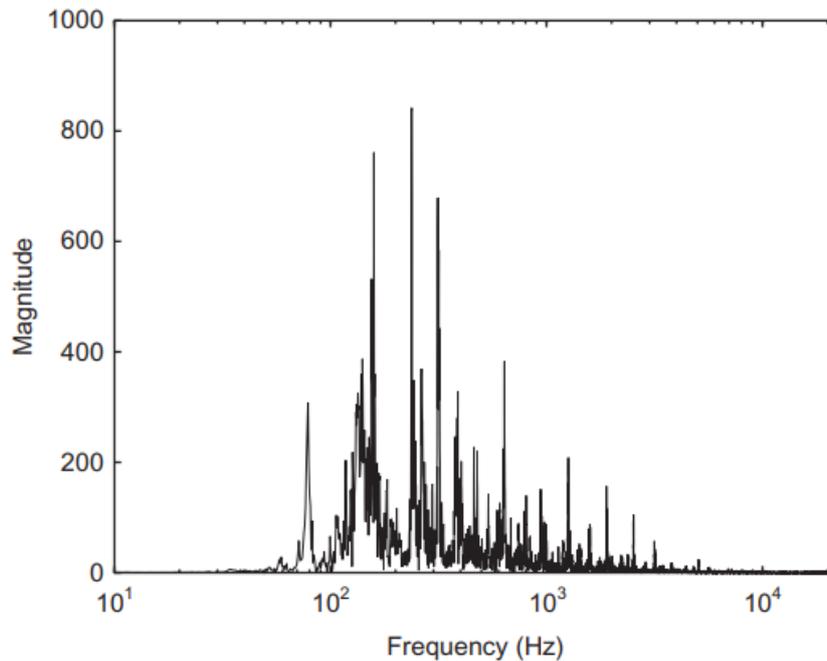


Figura 1.12: Spettro di ampiezza di un segnale [8]

Il passaggio inverso, dal dominio delle frequenze a quello del tempo, è comunque possibile grazie all'antitrasformata, o trasformata inversa, di Fourier:

$$s(t) = \mathcal{F}^{-1}\{S(f)\} = \int_{-\infty}^{+\infty} S(f)e^{j2\pi ft} df \quad (1.78)$$

Da un punto di vista puramente matematico, il dominio dello spettro di un segnale va da $-\infty$ a $+\infty$, cioè comprende sia le frequenze negative che quelle positive. D'altronde, nel caso dei segnali reali, è usuale rappresentare lo spettro nel solo dominio delle frequenze positive. Una spiegazione intuitiva è che la frequenza rappresenta il numero di oscillazioni nell'unità di tempo, perciò una quantità positiva. Una spiegazione più rigorosa si avvale della trasformata di Fourier: lo spettro di un segnale reale gode della proprietà di simmetria hermitiana, dal matematico francese Charles Hermite (1822-1901):

$$S(-f) = S^*(f) \quad (1.79)$$

dove l'asterisco denota il complesso coniugato. L'Eq. (1.79) afferma che la parte reale e la parte immaginaria dello spettro S sono, rispettivamente, una funzione pari

e una funzione dispari della frequenza. Un'affermazione equivalente è che lo spettro di ampiezza e lo spettro di fase sono, rispettivamente, una funzione pari e una funzione dispari della frequenza.

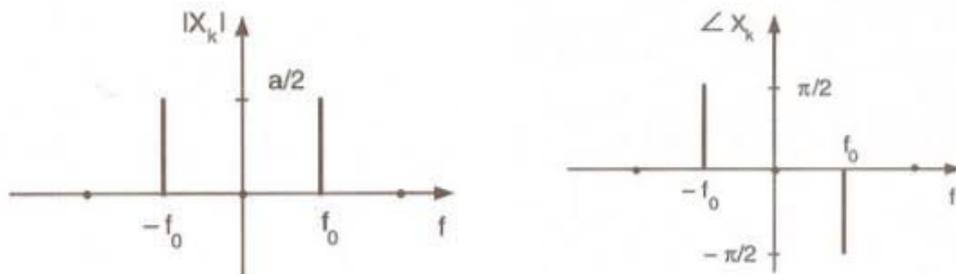


Figura 1.13: Simmetria hermitiana dello spettro di un segnale reale. A sinistra lo spettro di ampiezza, a destra quello di fase

L'informazione contenuta nel dominio delle frequenze negative è dunque una duplicazione dell'informazione nel dominio delle frequenze positive, motivo per cui non riusciamo a distinguere tra frequenze positive o negative. Per ottenere tale distinzione, si introduce un segnale $z(t)$, detto *segnale analitico*, il quale non presenta più la simmetria hermitiana:

$$Z(f) = \mathcal{F}\{z(t)\} = 0 \quad \forall f < 0 \quad (1.80)$$

Poiché un segnale reale è soggetto al vincolo di simmetria hermitiana, il segnale analitico non può essere reale, bensì complesso. Un primo beneficio di lavorare con il segnale analitico associato ad un segnale reale consiste nel dimezzare l'ampiezza di banda del segnale con la possibilità di dimezzare la minima frequenza di campionamento, oltre ad un dibattito sul fatto che le frequenze negative non abbiano una chiara interpretazione fisica. D'altra parte, in questa trattazione una frequenza è negativa in termini relativi e può essere concepita come una frequenza positiva che è stata traslata troppo a sinistra lungo l'asse delle frequenze dello spettro. Dati due segnali reali $s(t)$ e $y(t)$, un teorema stabilisce che il segnale complesso $z(t) = s(t) + y(t)$ è analitico se e solo se

$$Y(f) = -j \operatorname{sgn}(f) S(f) \quad (1.81)$$

dove $\operatorname{sgn}(f)$ è la funzione segno della frequenza. Assegnato un segnale reale $s(t)$, questo teorema ci consente di costruire il segnale analitico $z(t)$ associato ad esso, che d'ora in poi chiameremo, per brevità, *analitico associato*. La sua parte immaginaria, infatti, per il suddetto teorema è funzione del segnale reale secondo la seguente relazione:

$$y(t) = \Im\{z\} = \mathcal{F}^{-1}\{-j \operatorname{sgn}(f) \mathcal{F}\{s(t)\}\} \quad (1.82)$$

La combinazione di operazioni all'ultimo membro di questa catena di uguaglianze porta alla definizione di un operatore che prende il nome di trasformata integrale di Hilbert:

$$\mathcal{H}\{s(t)\} = \mathcal{F}^{-1}\{-j \operatorname{sgn}(f) \mathcal{F}\{s(t)\}\} \quad (1.83)$$

Quindi, dato un segnale reale $s(t)$, ottenere il suo analitico associato è immediato:

$$z(t) = s(t) + j\mathcal{H}\{s(t)\} \quad (1.84)$$

Il prodotto per l'unità immaginaria $j = e^{-j\pi/2}$ nella definizione di trasformata di Hilbert ci dice che quest'ultima introduce un ritardo di $\pi/2$ radianti, cioè produce un segnale cosiddetto *in quadratura* $Q(t)$ rispetto al segnale in ingresso, anche detto *in fase* e indicato con $I(t)$.

Nel nostro caso di interesse, l'analitico associato di un segnale sinusoidale è:

$$z(t) = a(t) \cos \theta(t) + j\mathcal{H}\{a(t) \cos \theta(t)\} = a(t) \cos \theta(t) + j a(t) \sin \theta(t) \quad (1.85)$$

Riconosciamo all'ultimo membro di questa catena di uguaglianze la formula di Eulero, quindi, per un segnale reale sinusoidale:

$$z(t) = a(t)e^{j\theta(t)} \quad (1.86)$$

Questo risultato è valido fintanto che la variazione di $a(t)$ nel tempo è sufficientemente lenta.

L'Eq. (1.86) fornisce una nuova rappresentazione del segnale: un vettore di modulo $a(t)$ che ruota nel piano complesso con velocità angolare $\omega = 2\pi f$, spazzando un angolo pari alla fase θ del segnale. Il segnale analitico porta dunque ad una nuova definizione dell'ampiezza e della fase di un segnale, cioè come modulo e argomento di un numero complesso. Il vantaggio di definire la fase come l'argomento dell'analitico associato è la possibilità di visualizzarla graficamente e ricavarla per via geometrica:

$$\theta(t) = \arctan \left[\frac{Q(t)}{I(t)} \right]$$

La definizione di frequenza rimane comunque quella dell'Eq. (1.75) [8].

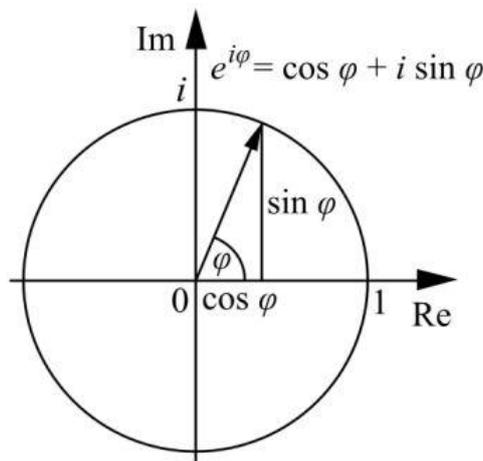


Figura 1.14: Rappresentazione di un numero complesso come vettore nel piano di Gauss. In questo esempio, il modulo è unitario e l'argomento, o fase, è l'angolo φ . Fonte: https://it.wikipedia.org/wiki/Formula_di_Eulero

1.2.2 Segnali analogici, segnali digitali e campionamento

I segnali finora trattati hanno in comune la caratteristica di essere funzioni continue del tempo. Segnali di questo tipo sono detti analogici. Esistono poi i segnali tempo-discreti,

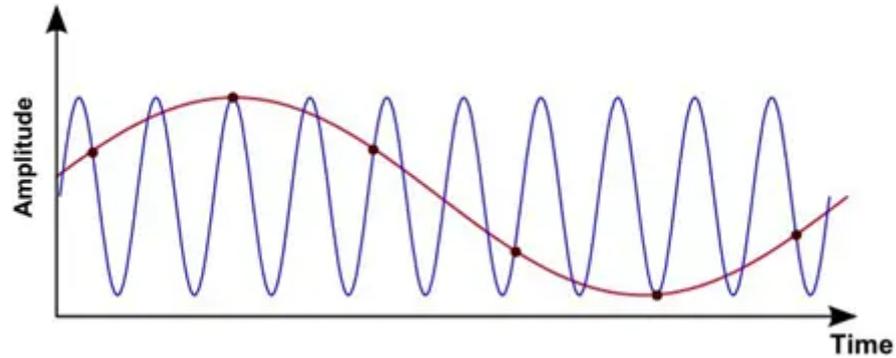


Figura 1.16: Aliasing di un segnale sinusoidale sotto-campionato. In blu è riportato il segnale originale, in rosso uno degli infiniti alias passante per i campioni rappresentati dai pallini neri. Fonte: <https://www.dataphysics.com/blog/dynamic-signal-analysis/dynamic-signal-analysis-review-part-3-aliasing/>

illustra una delle infinite possibili sinusoidi passanti per i campioni del segnale originale sotto-campionato.

C'è poi un'altra questione legata all'aliasing. Lo spettro di un segnale campionato è costituito dallo spettro del segnale analogico originale più altri spettri cosiddetti laterali, aventi la stessa ampiezza di banda dello spettro originale e centrati intorno a frequenze multiple della frequenza di campionamento. Se la massima frequenza del segnale è superiore alla frequenza di Nyquist f_N , cioè la metà della frequenza di campionamento, allora gli spettri laterali entrano nell'intervallo $[-f_N, f_N]$, e in alcuni casi si sovrappongono allo spettro originale, distorcendolo e impedendo così la corretta ricostruzione del segnale.

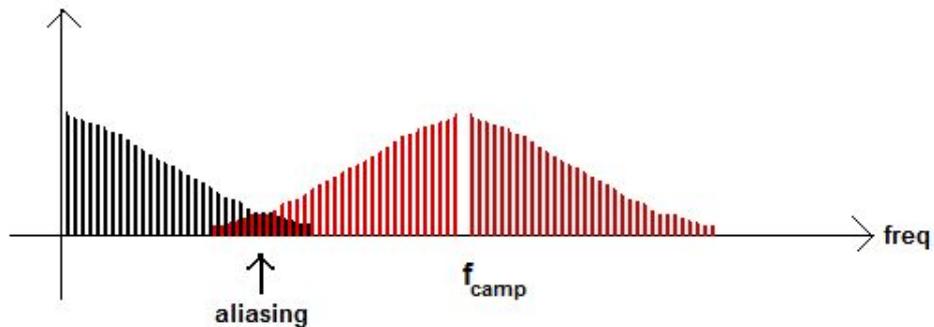


Figura 1.17: Aliasing di un segnale dal punto di vista spettrale. Lo spettro laterale, riportato in nero, si sovrappone allo spettro del segnale, indicato in rosso. Fonte: <http://www.robertopasini.com/index.php/2-uncategorised/313-hw-aliasing>

Capitolo 2

Phase-Locked Loops

Uno strumento ancora oggi impiegato per l'elaborazione dei segnali ai fini della determinazione d'orbita è il sistema PLL, acronimo di Phase-Locked Loop, o anello ad aggancio di fase, un circuito che insegue e aggancia la fase del segnale di riferimento in ingresso con lo scopo di generare in uscita un segnale che sia sincronizzato in frequenza, così come in fase, al segnale in ingresso. Per soddisfare tale funzione, il PLL è un sistema a catena chiusa, cioè presenta un ramo in retroazione e un comparatore. L'architettura di base del PLL consiste in tre elementi principali nel ramo in catena diretta e in una retroazione unitaria. I tre elementi principali sono il comparatore di fase, il filtro passa-basso e l'oscillatore controllato in tensione e il loro funzionamento sarà spiegato in dettaglio nel seguito.

Si dice che il PLL è nella condizione di aggancio o di lock quando la frequenza media del segnale in uscita eguaglia quella del segnale in ingresso, tuttavia questo non implica che l'errore di fase debba essere nullo. È bene comunque che tale errore si mantenga sufficientemente piccolo affinché il PLL non perda l'aggancio.

Nel tracciare il segnale di riferimento, il PLL svolge una seconda, ma non meno importante, azione, cioè filtrare il segnale in ingresso dal rumore per produrne una versione più pulita.

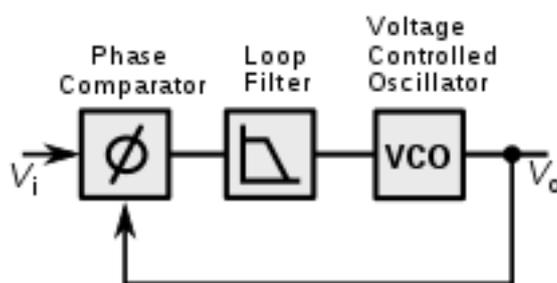


Figura 2.1: Diagramma a blocchi del circuito PLL

2.1 Caratteristiche principali

Nel progetto di un PLL, è importante definire la sua ampiezza di banda o Loop Bandwidth B_L , misurata in Hz. Sono state formulate diverse definizioni per tale parametro, ma,

essendo il comportamento del PLL assimilabile a quello di un filtro, opteremo per due definizioni inerenti alla teoria dei filtri e che si basano sul diagramma di Bode del modulo della funzione di risposta armonica $|H(j\omega)|$:

1. *3 dB bandwidth*: è l'ampiezza di banda in corrispondenza della quale $|H(j\omega)|$ è inferiore di 3 dB al suo valore massimo
2. *equivalent noise bandwidth*: è l'ampiezza di banda di un filtro rettangolare il cui diagramma di $|H(j\omega)|^2$ ha la stessa area sottesa dal diagramma di $|H(j\omega)|^2$ del filtro in esame

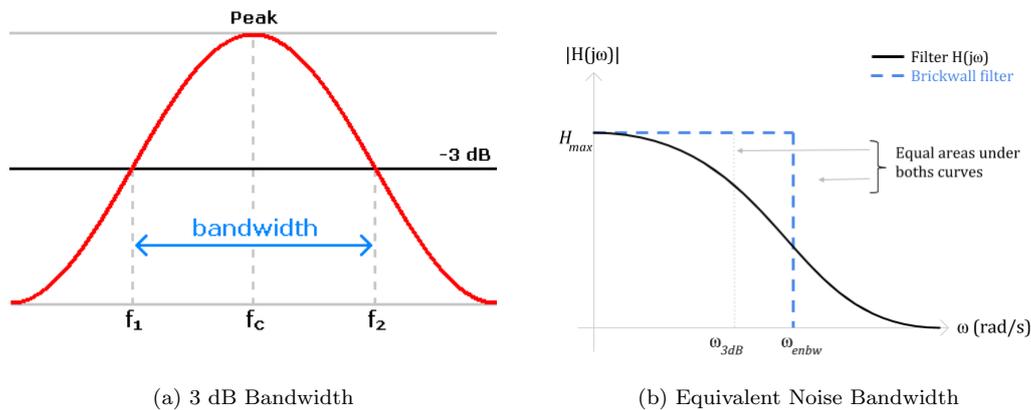


Figura 2.2: Definizioni di Loop Bandwidth.
 Fonte (a): <https://it.wikipedia.org/wiki/File:Bandwidth.png>.
 Fonte (b): <https://analog.intgckts.com/equivalent-noise-bandwidth/>

In particolare, la definizione di equivalent noise bandwidth è la più utile in quanto esprime la quantità di rumore che il filtro lascia passare.

La scelta dell'ampiezza di banda di un circuito PLL è il compromesso tra due requisiti contrastanti: una banda ampia facilita l'aggancio della fase, ma lascia passare una maggiore quantità di rumore; al contrario, una banda stretta lascia passare meno rumore, ma non garantisce che il PLL sia in grado di agganciare la fase, soprattutto se il segnale in ingresso ha una dinamica elevata, vale a dire una rapida variazione della frequenza.

Il circuito PLL è un sistema dinamico, cioè un sistema nel quale l'uscita in un dato istante di tempo dipende dall'andamento dell'ingresso a partire da un tempo iniziale fino a quell'istante, oltre che dalle condizioni iniziali. Per semplicità, considereremo il sistema PLL tempo-invariante, cioè tale che il suo comportamento non subisca variazioni con il passare del tempo.

In generale, il PLL è un sistema non-lineare, tuttavia, se l'errore di fase tra il segnale in ingresso e quello in uscita è molto piccolo, è possibile adottare una linearizzazione in un intorno dell'errore di fase nullo. Sotto questa ipotesi, nelle seguenti sezioni linearizzeremo i diversi componenti del PLL per ottenere un modello che possa essere studiato con i potenti strumenti dell'analisi dei sistemi lineari, in particolare la trasformata di Laplace e le funzioni di trasferimento.

2.2 Phase Detector

Il primo elemento della catena diretta è il comparatore di fase, o Phase Detector PD, un circuito che riceve in ingresso il segnale di riferimento e quello generato dall'oscillatore al termine della catena diretta per fornire in uscita un segnale di tensione funzione dell'errore di fase tra i due segnali. Esistono svariate tipologie di comparatori di fase, ma quello di nostro interesse è il primo ad essere stato utilizzato nella storia, cioè il phase detector moltiplicatore. Definiamo il segnale sinusoidale in ingresso (per semplicità, trascureremo il rumore additivo):

$$u_1 = U_1 \sin(\omega_1 t + \theta_{o,1}) \quad (2.1)$$

essendo $\omega = 2\pi f$ la pulsazione e θ_o la fase iniziale. Il secondo segnale in ingresso è la senoide generata dall'oscillatore controllato in tensione e proveniente dal ramo in retroazione:

$$u_2 = U_2 \cos(\omega_2 t + \theta_{o,2}) \quad (2.2)$$

Il segnale di tensione in uscita dal phase detector è dunque pari al prodotto dei due segnali in ingresso:

$$u_d = U_1 U_2 \sin(\omega_1 t + \theta_{o,1}) \cos(\omega_2 t + \theta_{o,2}) \quad (2.3)$$

Se il PLL è nella condizione di lock, allora, grazie ad una delle formule trigonometriche di Werner, l'Eq. (2.3) si semplifica:

$$\begin{aligned} u_d &= U_1 U_2 \sin(\omega_1 t + \theta_{o,1}) \cos(\omega_2 t + \theta_{o,2}) = \\ &= \frac{U_1 U_2}{2} [\sin(2\omega t + \theta_{o,1} + \theta_{o,2}) + \sin(\theta_e)] = \\ &= K_{PD} [\sin(2\omega t + \theta_{o,1} + \theta_{o,2}) + \sin(\theta_e)] \end{aligned} \quad (2.4)$$

dove $\theta_e = \theta_1 - \theta_2$ è l'errore di fase, ossia la differenza tra la fase del segnale in ingresso e di quello in uscita. La costante K_{PD} è il cosiddetto guadagno del comparatore di fase. Il prodotto dei due segnali in ingresso al moltiplicatore è dunque la somma di un contributo costante funzione dell'errore di fase e di un contributo sinusoidale con frequenza pari al doppio della frequenza in ingresso (ricordiamo che quando il PLL è in lock il segnale in ingresso e quello in retroazione hanno la stessa frequenza media). Poiché quest'ultimo contributo è indesiderato, il segnale di tensione deve passare attraverso il successivo elemento della catena diretta: il filtro passa-basso. Se poi l'errore di fase è sufficientemente piccolo, è possibile eseguire una linearizzazione. Trascurando il termine sinusoidale dell'Eq. (2.4), poiché soppresso dal successivo filtro passa-basso, si ha che:

$$u_d = K_{PD} \theta_e \quad (2.5)$$

Tale approssimazione lineare ci consente di definire la funzione di trasferimento del phase detector. Applicando la trasformata di Laplace, l'Eq. (2.5) diventa:

$$U_d(s) = K_{PD} \Theta_e(s) \quad (2.6)$$

dove s è la variabile complessa nel dominio della trasformata di Laplace. La funzione di trasferimento del phase detector, quindi, è:

$$H_{PD} = \frac{U_d(s)}{\Theta_e(s)} = K_{PD} \quad (2.7)$$

ossia un semplice guadagno.

2.3 Lowpass Filter

Il segnale di tensione in uscita dal comparatore di fase accede al successivo elemento del PLL, cioè il filtro passa-basso, anche detto Loop Filter LF, per l'eliminazione del contributo sinusoidale di cui si è discusso in precedenza. Esistono numerosi filtri, ma nell'ambito di questa trattazione ne considereremo solo due. Il primo filtro passa-basso è un elemento proporzionale-integrale, la cui funzione di trasferimento è la seguente:

$$H_{LF} = K_P + \frac{K_I}{s} \quad (2.8)$$

dove K_P è il guadagno della parte proporzionale e K_I il guadagno di quella integrale. Il secondo filtro che considereremo è in generale un elemento proporzionale-integrale-derivativo. Nell'ambito di questa trattazione:

$$H_{LF} = K_P + \frac{K_{I1}}{s} + \frac{K_{I2}}{s^2} \quad (2.9)$$

2.4 Voltage-Controlled Oscillator

Una volta filtrato, il segnale di tensione funzione dell'errore di fase entra nel terzo elemento del PLL, l'oscillatore controllato in tensione, o Voltage-Controlled Oscillator VCO. L'oscillatore è un piccolo circuito che produce un segnale sinusoidale e che modifica la propria frequenza sulla base dell'errore di fase in ingresso al fine di minimizzare l'errore di fase stesso. L'equazione fondamentale del VCO è la seguente:

$$\omega_2 = \omega_1 + \Delta\omega = \omega_1 + K_{VCO}u_f \quad (2.10)$$

essendo K_{VCO} il guadagno dell'oscillatore e u_f il segnale di errore di fase in uscita dal filtro.

Poiché la pulsazione ω è la derivata nel tempo della fase in uscita θ_2 , segue che:

$$\theta_2 = \int \Delta\omega dt = K_{VCO} \int u_f dt \quad (2.11)$$

Applicando la trasformata di Laplace, si ottiene la funzione di trasferimento dell'oscillatore:

$$\Theta_2 = \frac{K_{VCO}}{s} U_f \Rightarrow H_{VCO} = \frac{K_{VCO}}{s} \quad (2.12)$$

Come ci aspettavamo dall'Eq. (2.11), la funzione di trasferimento dell'oscillatore è quella di un integratore [10][11].

2.5 Analisi lineare

Avendo a disposizione le funzioni di trasferimento dei singoli elementi, si può finalmente ricavare la funzione di trasferimento dell'intero sistema retro-azionato. Per la regola di Mason:

$$H_{PLL}(s) = \frac{G(s)}{1 + G(s)} \quad (2.13)$$

dove $G(s)$ è la funzione di trasferimento del ramo in catena diretta:

$$G(s) = K_{PD} \frac{K_{VCO}}{s} H_{LF} \quad (2.14)$$

Sostituendo l'Eq. (2.14) nella (2.13), si giunge a:

$$H_{PLL}(s) = \frac{K_{PD} K_{VCO} H_{LF}}{s + K_{PD} K_{VCO} H_{LF}} \quad (2.15)$$

Se il filtro passa-basso è un proporzionale-integrale, l'Eq. (2.15) diventa:

$$H_{PLL}(s) = \frac{K_{PD} K_{VCO} K_P s + K_{PD} K_{VCO} K_I}{s^2 + K_{PD} K_{VCO} K_P s + K_{PD} K_{VCO} K_I} = \frac{K_1 s + K_2}{s^2 + K_1 s + K_2} \quad (2.16)$$

Poiché il denominatore della funzione di trasferimento è un polinomio di secondo grado nella variabile complessa s , tale PLL è detto del secondo ordine.

Nel caso in cui il filtro fosse, invece, un proporzionale-integrale-derivativo, il PLL è del terzo ordine:

$$H_{PLL}(s) = \frac{K_1 s^2 + K_2 s + K_3}{s^3 + K_1 s^2 + K_2 s + K_3} \quad (2.17)$$

dove:

$$\begin{aligned} K_1 &= K_{PD} K_{VCO} K_P \\ K_2 &= K_{PD} K_{VCO} K_{I1} \\ K_3 &= K_{PD} K_{VCO} K_{I2} \end{aligned} \quad (2.18)$$

Sottolineiamo ancora una volta che tali risultati sono validi solamente quando il PLL è approssimato ad un sistema dinamico lineare e tempo-invariante. Questo implica delle considerazioni sull'ingresso e l'uscita. Si considerino le Eq. (2.6) e (2.12):

1. l'uscita del PLL lineare non è un segnale sinusoidale, bensì la sua fase θ_2
2. l'uscita del phase detector è semplicemente una costante che moltiplica l'errore di fase θ_e , per cui l'ingresso del PLL lineare non è il segnale sinusoidale che vogliamo tracciare, ma la sua fase θ_1

In conclusione, il PLL lineare non elabora il segnale, ma la sua fase.

Grazie alle funzioni di trasferimento, è possibile studiare l'errore di fase a regime del PLL lineare. Sulla base del diagramma a blocchi in Fig. (2.1), l'errore di fase nel dominio della trasformata di Laplace è pari a:

$$E(s) = \frac{1}{1 + G(s)} \Theta_1(s) \quad (2.19)$$

dove $G(s)$ è sempre la funzione di trasferimento del ramo in catena diretta e $\Theta_1(s)$ è la fase in ingresso. Per mezzo del teorema del valore finale della trasformata di Laplace, l'errore di fase a regime è il seguente limite:

$$E_\infty(s) = \lim_{s \rightarrow 0} sE(s) \quad (2.20)$$

Antitrasformando si ottiene infine l'errore a regime nel dominio del tempo.

La tabella (2.1) mostra gli errori di fase a regime per tre diverse dinamiche di frequenza, sia nel caso di PLL del secondo ordine che in quello di PLL del terzo ordine. Osserviamo che, all'aumentare della dinamica del segnale, le prestazioni di un PLL del terzo ordine sono superiori a quelle di un PLL del secondo ordine.

ordine del PLL	f costante	f lineare	f parabolica
2	0	$\frac{2\pi\alpha}{K_2}$	∞
3	0	0	$\frac{2\pi\beta}{K_3}$

Tabella 2.1: Errore di fase a regime per tre diverse dinamiche di frequenza del segnale. Le costanti α e β sono, rispettivamente, il rateo temporale e l'accelerazione della frequenza

Come è stato ribadito più volte, il modello di PLL lineare è valido fintanto che questo si trova nello stato di lock. All'infuori dell'aggancio il comportamento di un PLL non può essere linearizzato e si entra nell'analisi non-lineare. Sia $|\Delta f|$ la differenza in valore assoluto tra la frequenza del segnale in ingresso e quella del segnale in uscita. In base all'entità di questa differenza si distinguono i seguenti range di funzionamento del PLL nella condizione di unlock [11]:

1. hold-in range Δf_H : è la massima differenza di frequenza per la quale il PLL è capace di ritornare allo stato di lock in seguito ad un disturbo di entità opportuna. La definizione equivale a quella di stabilità statica
2. pull-in range Δf_P , anche detto acquisition range o capture range: è la massima differenza di frequenza per la quale il PLL, dopo aver perso l'aggancio, raggiunge la condizione di lock in seguito ad un transitorio, motivo per cui tale definizione equivale a quella di equilibrio dinamico

- lock-in range Δf_L : è la massima differenza di frequenza per cui il PLL è in grado di ritornare alla condizione di lock in seguito ad una brusca variazione della frequenza di riferimento, senza che vi sia *cycle slipping*, cioè un salto di fase superiore a 2π radianti.

In generale, l'ampiezza dei tre range è tale da rispettare la seguente relazione:

$$\Delta f_L < \Delta f_P < \Delta f_H$$

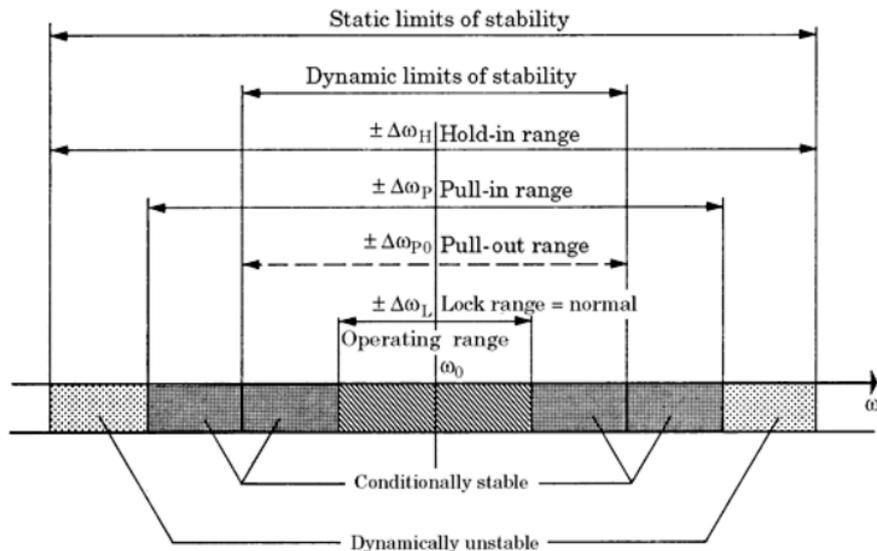


Figura 2.3: Range di funzionamento del PLL all'infuori dello stato di lock [11]

2.6 Digital Phase-Locked Loops

L'analisi svolta fino a questo punto riguarda i sistemi analogici, cioè tempo-continui. Come tanti altri dispositivi, anche i PLL si sono evoluti dalla forma analogica a quella digitale. Le ragioni principali sono le seguenti: costi minori, maggiore semplicità nella realizzazione, assenza di problemi presenti nei sistemi analogici e possibilità di immagazzinare i segnali, operazione impraticabile per sistemi analogici.

A differenza dei sistemi analogici, numerosi DPLL, acronimo di Digital Phase-Locked Loop, presentano grosse non-linearità che rendono difficile, se non impossibile, una qualunque approssimazione, anche nel caso di piccoli errori di fase. Tuttavia, anche in assenza di grosse non-linearità, ogni DPLL soffre degli effetti della quantizzazione: la quantizzazione è un'operazione non lineare le cui conseguenze sono maggiormente significative per piccoli errori di fase. Onde evitare le complicazioni dell'analisi non-lineare, si considera una quantizzazione abbastanza fine da poterla trascurare in un'approssimazione del primo ordine. Così facendo, anche il DPLL può essere studiato con l'analisi dei sistemi lineari [10].

Poiché i sistemi digitali sono tempo-discreti, è necessario passare dal dominio continuo della trasformata di Laplace a quello discreto. Tale operazione è svolta dalla trasformata Z. Nella nostra trattazione, utilizzeremo la trasformata Zeta secondo l'approssimazione di Tustin:

$$s = \frac{2z - 1}{t_s z + 1} \quad (2.21)$$

essendo s la variabile complessa nel dominio continuo della trasformata di Laplace, t_s il periodo di campionamento e z la variabile complessa nel dominio discreto. La trasformata di Tustin fornisce il miglior match nel dominio delle frequenze tra la forma tempo-continua e quella tempo-discreta del sistema, oltre a conservare la stabilità del sistema.

2.7 Implementazione in Matlab e Simulink

Come constatato in precedenza, un PLL lineare può ricevere in ingresso solo la fase del segnale di riferimento e fornire solo la fase del segnale in uscita. Solitamente, nell'ambito della radio scienza, i segnali si presentano nella loro forma analitica, cioè segnali complessi con una parte in fase e una in quadratura, per cui la fase è ottenibile dall'applicazione dell'arcotangente a quattro quadranti $\arctan_2(\cdot)$. L'immagine della funzione $\arctan_2(\cdot)$ è l'intervallo $[-\pi, \pi]$: la fase non è continua, ma ogni volta che supera i π radianti torna indietro di un ciclo, vale a dire 2π radianti. Si parla quindi di *wrapped phase*.

Il codice implementato in Matlab durante il tirocinio simulava la risposta delle funzioni di trasferimento dell'intero sistema retro-azionato, corrispondenti a quelle riportate nell'Eq. (2.16) e (2.17), e la simulazione avveniva per mezzo dei comandi `tf` e `lsim`, abbreviazioni di *transfer function* e *linear simulator*. Quest'ultimo comando non accetta in ingresso grandezze discontinue, come appunto la fase ricavata dall'arcotangente, motivo per cui la fase doveva essere "srotolata" tramite il comando `unwrap`: ogni volta che il salto di fase tra due angoli consecutivi è maggiore o uguale a π radianti, la fase è traslata di $\pm 2\pi$ radianti finché il salto non è inferiore a π radianti.

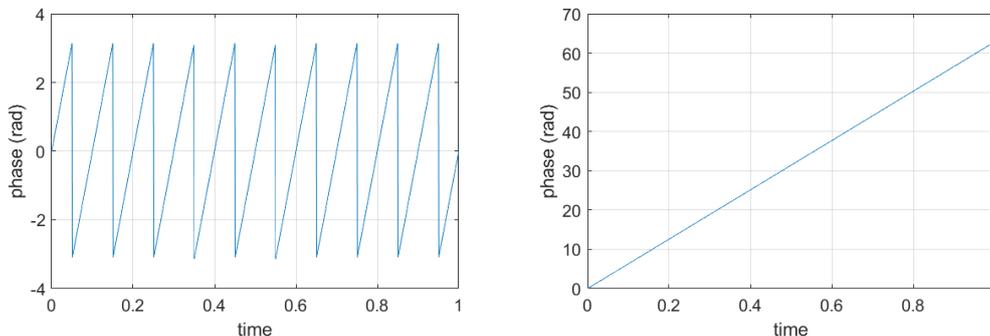


Figura 2.4: A sinistra la fase di un segnale sinusoidale con frequenza di 10 Hz "avvolta" tra $-\pi$ e π radianti; a destra, la stessa fase dopo l'unwrap

Il problema è che quando il rumore diventa importante le oscillazioni della fase sono tali da "ingannare" l'unwrap, il quale cessa di agire correttamente. Durante il tirocinio è stato osservato che l'unwrap funzionava fino a $CN0 = 40$ dB-Hz, tuttavia era atteso un funzionamento del PLL fino a $CN0 = 15$ dB-Hz.

Poiché il PLL non si limita a tracciare il segnale in ingresso, ma lo filtra anche dal rumore, si è pensato di implementare un PLL che riceva e fornisca un segnale anziché una fase. In questo modo è possibile spostare l'operazione di unwrap della fase da monte a valle della simulazione e, avendo in uscita una versione più pulita del segnale di riferimento,

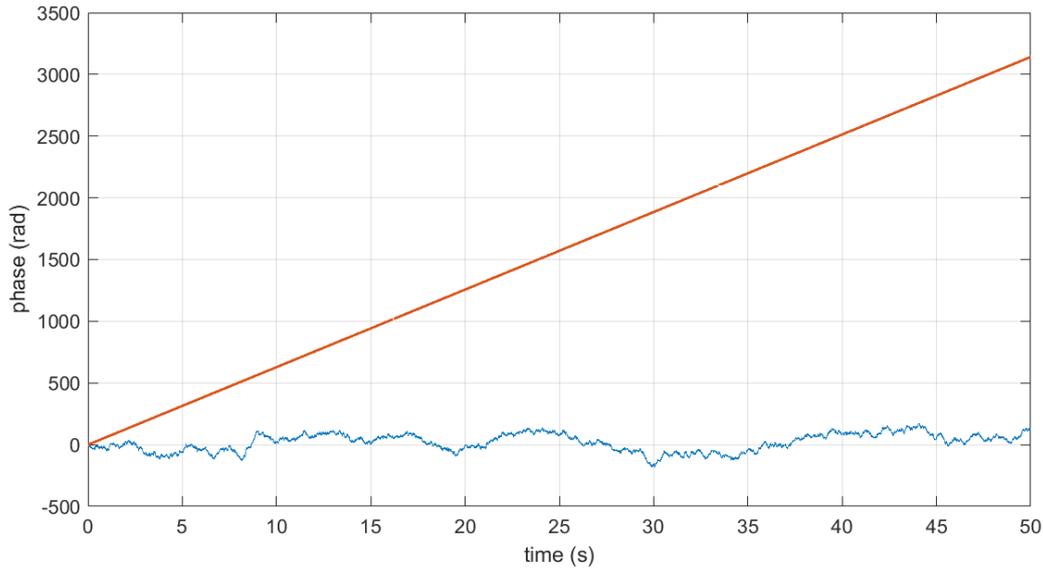


Figura 2.5: Unwrap scorretto da parte di Matlab della fase di un segnale sinusoidale con frequenza di 10 Hz e $CN0 = 20$ dB-Hz. In azzurro l'unwrap di fase di Matlab, in arancione quello che dovrebbe essere il corretto risultato dell'unwrap

l'unwrap della fase e la conseguente ricostruzione della frequenza del segnale saranno più agevoli anche per $CN0$ più bassi. Il nuovo modello di PLL è stato implementato in Simulink sulla base dei modelli di PLL proposti dal Library Browser di Simulink.

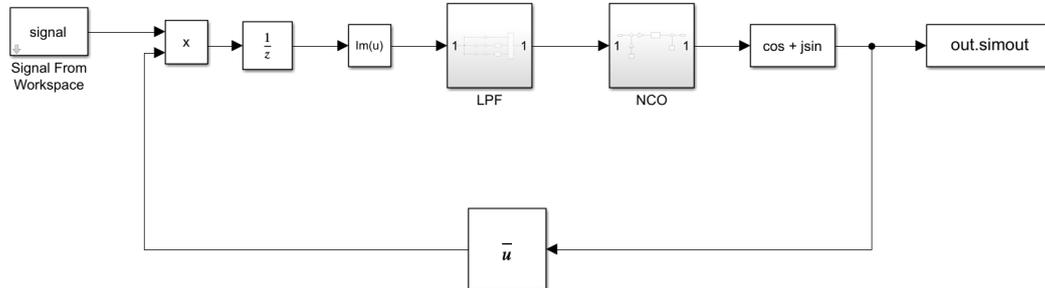


Figura 2.6: Diagramma a blocchi del PLL implementato su Simulink

Il comparatore di fase consiste in un blocco moltiplicatore che esegue il prodotto tra il segnale analitico in ingresso dal Workspace di Matlab e il complesso coniugato del segnale analitico in uscita dall'oscillatore:

$$\begin{aligned}
 u_d &= A_{in} e^{j\theta_{in}} e^{-j\theta_{out}} = A_{in} e^{j(\theta_{in} - \theta_{out})} = \\
 &= A_{in} [\cos(\theta_{in} - \theta_{out}) + j \sin(\theta_{in} - \theta_{out})]
 \end{aligned}
 \tag{2.22}$$

dove A_{in} è l'ampiezza del segnale in ingresso. In tal modo, l'argomento del prodotto è la differenza di fase tra i due segnali. Ponendo che il PLL sia nello stato di lock (vedremo a breve come rendere vera quest'assunzione), l'errore di fase è sufficientemente piccolo per approssimare la parte immaginaria del prodotto dei due segnali nell'Eq. (2.22) con l'errore di fase stesso. In altre parole, il PLL implementato è ancora lineare. Questo

spiega la presenza del blocco $\text{Im}(u)$ a valle del moltiplicatore: esso estrae la sola parte immaginaria del prodotto, pari all'errore di fase nell'approssimazione lineare.

Il blocco di unit delay posto prima dell'estrazione della parte immaginaria del prodotto dei due segnali è necessario per eliminare il loop algebrico e le problematiche annesse nel modello in Simulink. Un loop algebrico ha luogo quando uno degli ingressi ad un blocco dipende dall'uscita di quello stesso blocco attraverso un percorso diretto di retroazione. Per percorso diretto si intende ciascuno dei seguenti casi:

- funzione matematica
- guadagno
- prodotto
- spazio degli stati con matrice D non nulla
- somma
- funzione di trasferimento impropria, cioè quando il grado del numeratore è lo stesso del denominatore

In presenza di loop algebrici, Simulink utilizza un solutore non lineare per risolvere ad ogni iterazione il loop algebrico, ma al prezzo di un maggiore costo computazionale. Inoltre, in alcuni casi, il solutore potrebbe non essere in grado di risolvere il loop algebrico. Per queste ragioni, è stato necessario introdurre un blocco di unit delay: grazie ad esso, il ramo di retroazione non è più diretto, nel senso indicato sopra, e il loop algebrico è stato così rimosso.

Una volta estratto l'errore di fase, questo passa al blocco del filtro passa-basso.

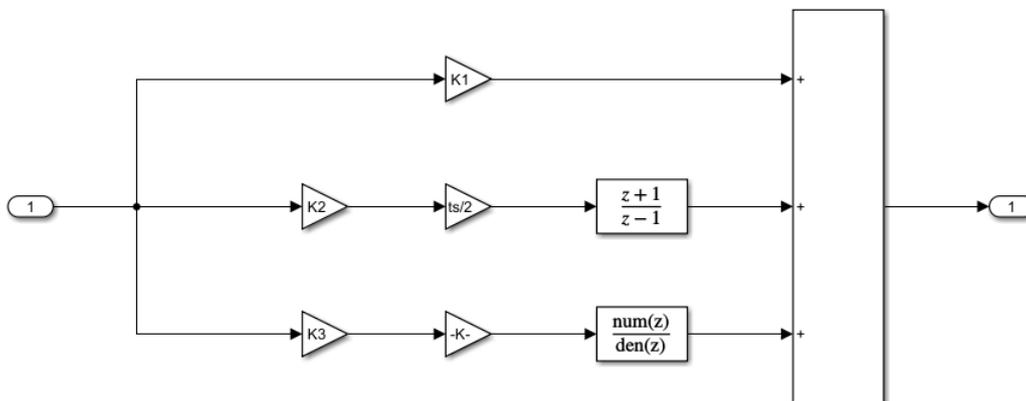


Figura 2.7: Diagramma a blocchi del Lowpass Filter LPF

Sfruttando la linearità del sistema, il filtro consiste nella sovrapposizione di tre rami:

1. proporzionale con guadagno K_P
2. integrale con guadagno K_{I1}
3. doppio integrale con guadagno K_{I2} , non nullo solo per PLL del terzo ordine

In particolare, essendo il segnale in ingresso campionato, il blocco integratore è stato convertito dalla forma tempo-continua a quella tempo-discreta attraverso la trasformata di Tustin, come svolto dal Jet Propulsion Laboratory della NASA e descritto in [12]:

$$\frac{1}{s} = \frac{t_s}{2} \frac{z + 1}{z - 1} \quad (2.23)$$

Il segnale di errore di fase in uscita dal filtro accede dunque all'oscillatore. Poiché il sistema è tempo-discreto, esso è detto Numerically-Controlled Oscillator o NCO.

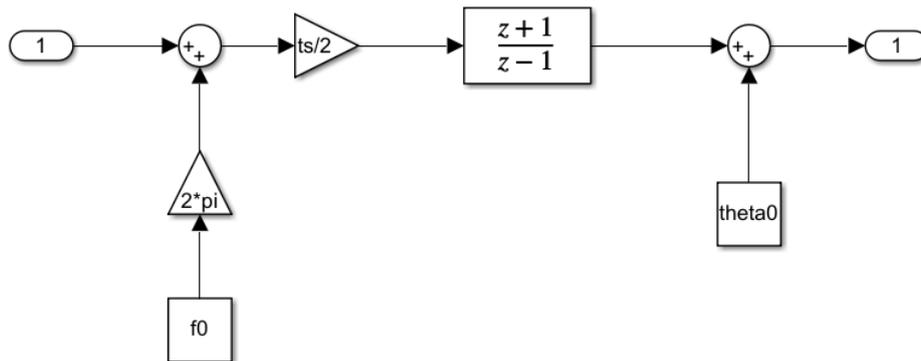


Figura 2.8: Diagramma a blocchi del Numerically-Controlled Oscillator NCO

Qui il segnale di controllo è sommato alla frequenza base f_0 dell'oscillatore, a sua volta moltiplicata per 2π al fine di ottenere una pulsazione, prima di passare al blocco integratore tempo-discreto. La fase in uscita dall'integratore è sommata alla fase iniziale del segnale di riferimento, dopodiché è utilizzata per generare il segnale analitico in uscita dall'oscillatore per mezzo del blocco $\cos + j\sin$. Il segnale in uscita dal NCO è poi trasferito al Workspace di Matlab per l'estrazione della fase e la ricostruzione della frequenza media.

Tornando alla frequenza base dell'oscillatore, essa è l'elemento che consente al PLL di entrare subito nello stato di lock e che rende quindi il suo comportamento lineare. Perché questo accada, la frequenza base dell'oscillatore deve essere pari alla frequenza iniziale del segnale in ingresso. Non essendo quest'ultima nota a priori, si è fatto uso dell'operatore periodogramma, ossia uno stimatore dello spettro di potenza del segnale. Il massimo dello spettro così stimato si ha in corrispondenza di quella che è una stima della frequenza iniziale del segnale. Tale stima è dunque assegnata all'oscillatore come frequenza base.

Un'altra condizione iniziale da specificare è la fase iniziale del segnale. La strategia adottata per la stima di questo dato è la seguente:

1. eseguire la Fast Fourier Transform FFT del segnale in ingresso
2. calcolare lo spettro di fase tramite la function `angle` la quale restituisce l'argomento di un numero complesso nell'intervallo $[-\pi, \pi]$ (ricordiamo che la trasformata di Fourier è in generale una quantità complessa)
3. scegliere la fase che corrisponde alla stima di frequenza iniziale fornita dal periodogramma

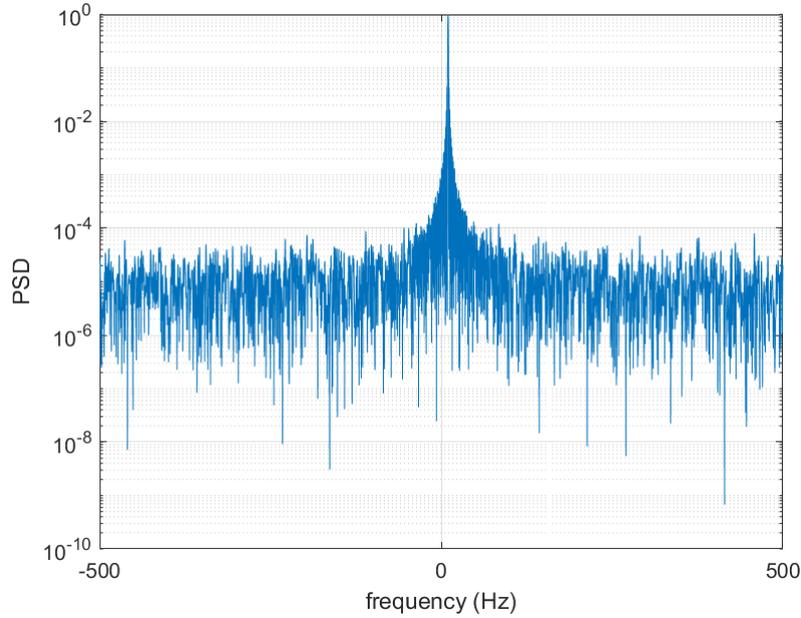


Figura 2.9: Periodogramma di un segnale sinusoidale con frequenza costante di 10 Hz e $CN0 = 50$ dB-Hz. PSD è l'acronimo di Power Spectral Density

Rimane la questione della scelta dei guadagni che compaiono nel modello implementato. Tale scelta è basata sui coefficienti K_1 , K_2 e K_3 delle funzioni di trasferimento (2.16) e (2.17) dell'intero PLL lineare. La tabella (2.2) illustra i coefficienti K_1 , K_2 e K_3 che sono stati utilizzati. Essi sono proposti dal Jet Propulsion Laboratory della NASA nel suo documento [6]. In questa tabella, i coefficienti determinano la natura del PLL, cioè se questo è sotto-smorzato o sovra-smorzato, sia nel caso di PLL del secondo ordine che in quello di PLL del terzo ordine. Osserviamo inoltre che tali coefficienti sono espressi in funzione della Loop Bandwidth B_L .

Facendo riferimento alle Eq. (2.16), (2.17) e (2.18), sono possibili molteplici combinazioni dei guadagni. Per l'implementazione è stata scelta la seguente combinazione:

$$\begin{aligned}
 K_{PD} &= 1 \\
 K_{VCO} &= 1 \\
 K_P &= K_1 \\
 K_{I1} &= K_2 \\
 K_{I2} &= K_3
 \end{aligned} \tag{2.24}$$

ordine del PLL	smorzamento	K_1	K_2	K_3
2	sotto-smorzato	$\frac{8}{3}B_L$	$\frac{1}{2}K_1^2$	-
	sovra-smorzato	$\frac{16}{5}B_L$	$\frac{1}{4}K_1^2$	-
3	sotto-smorzato	$\frac{60}{23}B_L$	$\frac{4}{9}K_1^2$	$\frac{2}{27}K_1^3$
	sovra-smorzato	$\frac{32}{11}B_L$	$\frac{1}{3}K_1^2$	$\frac{1}{27}K_1^3$

Tabella 2.2: Lista dei coefficienti K_1 , K_2 e K_3 proposti dal JPL della NASA

2.8 Verifica del modello

Una volta che il modello del PLL è stato implementato in Simulink, occorre testarne la validità. Una prima verifica è quella dell'errore di fase a regime.

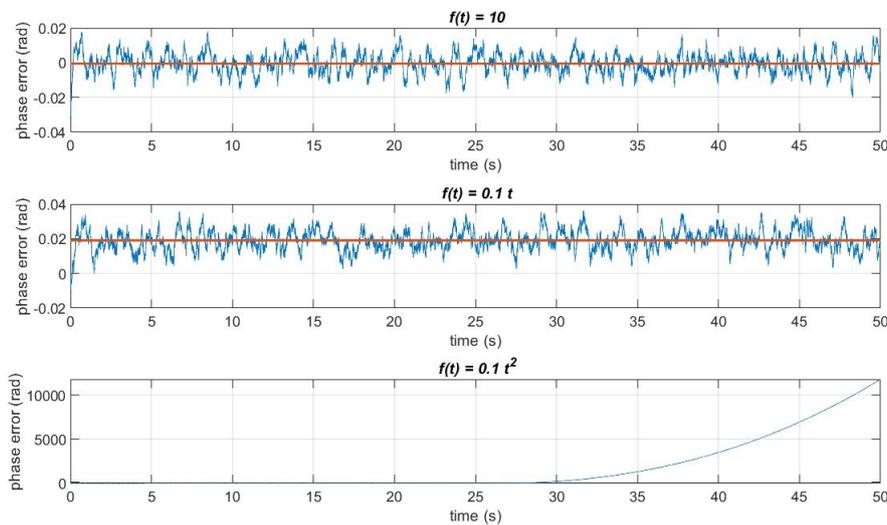


Figura 2.10: Errore di fase nel tempo con PLL del secondo ordine. La linea arancione è la media dei campioni

Entrambe le Fig. (2.10) e (2.11) sono state ottenute con $B_L = 3$ Hz, $T_c = 1$ s e $CN0 = 50$ dB-Hz e confermano gli andamenti a regime dell'errore di fase mostrati nella tabella (2.1): per il PLL del secondo ordine, l'errore di fase a regime è nullo per frequenza costante, costante e non nullo per frequenza lineare e divergente per frequenza parabolica;

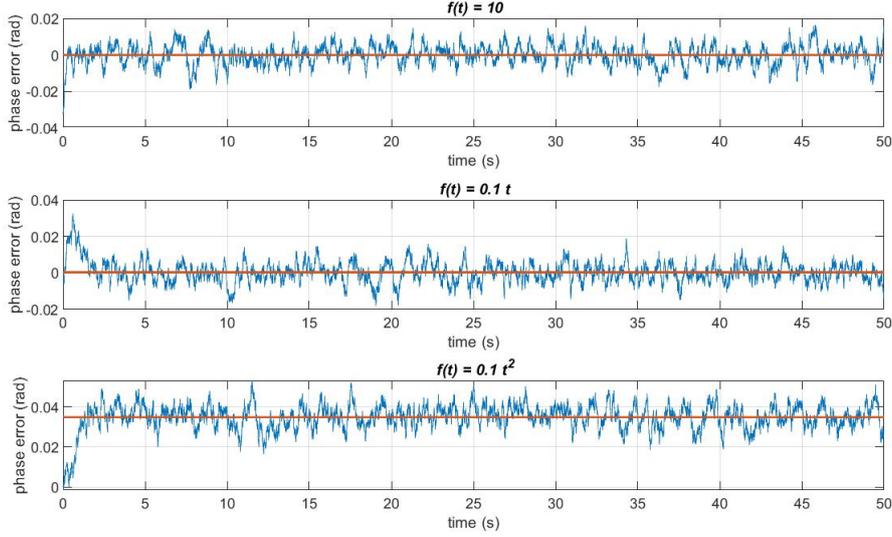


Figura 2.11: Errore di fase nel tempo con PLL del terzo ordine. La linea arancione è la media dei campioni

per il PLL del terzo ordine, l'errore di fase a regime è nullo sia per frequenza costante che per frequenza lineare e costante e non nullo per frequenza parabolica.

Successivamente è stato operato un confronto della varianza della stima di frequenza media con il limite inferiore di Cramer-Rao del PLL, ossia la minima varianza teorica raggiungibile. Ricordiamo che, essendo il segnale campionato, lo sarà anche la sua fase, perciò non è possibile calcolare la frequenza tramite l'Eq. (1.75), ma la si può calcolare come frequenza media su un intervallo di tempo, detto *count time* T_c :

$$f_m = \frac{1}{2\pi} \frac{\theta_b - \theta_a}{T_c} \quad (2.25)$$

dove θ_a e θ_b sono il primo e l'ultimo campione di fase in un count time.

Il JPL propone in [13] il seguente modello semplificato di limite inferiore di Cramer-Rao del PLL:

$$\sigma_f^2 = \frac{B_L}{2\pi^2 T_c^2 P_c / N_o} \quad (2.26)$$

dove σ_f^2 è la varianza nella stima di frequenza, B_L è la Loop Bandwidth e P_c/N_o è il carrier-to-noise density ratio.

La Fig. (2.12) mostra il risultato delle simulazioni: il nuovo modello implementato in Simulink è valido e migliore di quello implementato durante il tirocinio per CN0 fino al valore richiesto di 15 dB-Hz e in alcuni casi anche fino a 10 dB-Hz. Le dinamiche di frequenza analizzate sono le stesse che compaiono nelle Fig. (2.10) e (2.11).

Le simulazioni sono state effettuate su un numero quanto più ampio possibile di campioni di fase in modo da effettuare tante prove di stima della frequenza media. Maggiore è il numero di prove, più la varianza calcolata converge al valore atteso, cioè il limite inferiore di Cramer-Rao.

Le simulazioni con segnali a frequenza lineare e parabolica hanno dato risultati pressoché uguali al grafico a destra in Fig. (2.12).

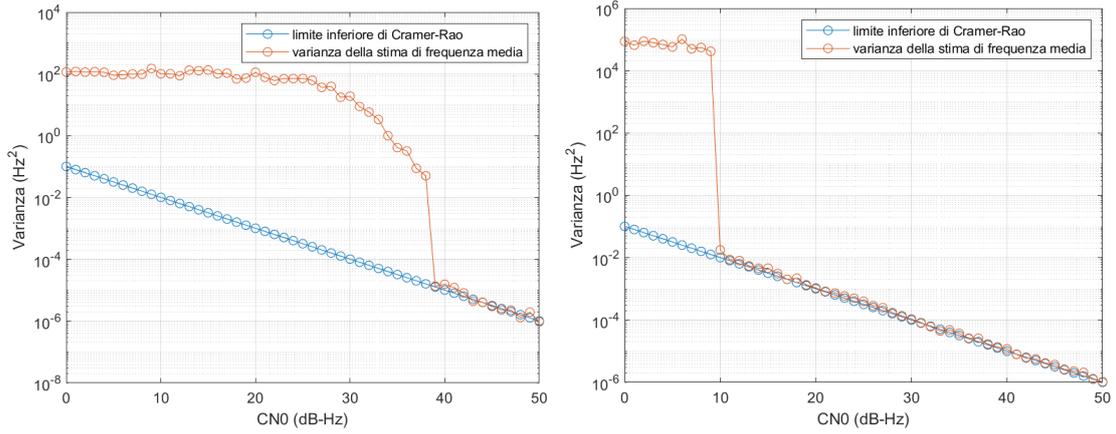


Figura 2.12: Confronto con il limite inferiore di Cramer-Rao del PLL al variare del CN0. A sinistra, il risultato ottenuto in tirocinio; a destra, quello del nuovo modello implementato in Simulink. Entrambi i grafici sono stati ottenuti con un segnale di durata $T = 500$ s a frequenza costante di 10 Hz, banda di loop $B_L = 2$ Hz e count time $T_c = 1$ s

Inoltre, aumentando la banda di loop, è stata osservato che aumentava il minimo CN0 per cui la stima di frequenza media seguiva il limite inferiore di Cramer-Rao del PLL. Una maggiore banda di loop, infatti, nonostante consenta di seguire meglio la dinamica di un segnale, lascia passare più rumore con ripercussioni sulla stima della frequenza media.

La stessa verifica è stata poi fatta scegliendo la Loop Bandwidth ideale indicata in [14], cioè tale da minimizzare il rumore senza perdere l'aggancio di fase, pari a:

$$B_{L,opt} = \frac{1}{2T_c} \quad (2.27)$$

essendo T_c il count time per la stima di frequenza media.

Ad esempio, se $T_c = 1$ s, allora la banda di loop ideale è pari a 0.5 Hz.

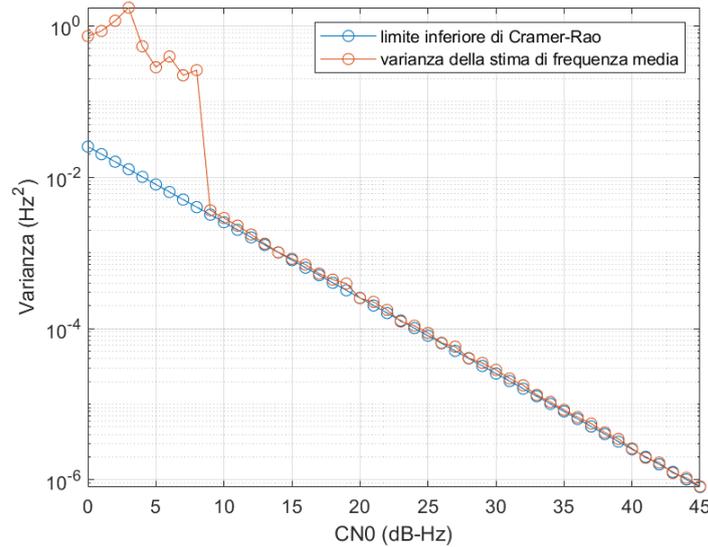


Figura 2.13: Confronto con il limite inferiore di Cramer-Rao del PLL per segnale di durata $T = 500$ s a frequenza costante di 10 Hz, $T_c = 1$ s e $B_L = B_{L,opt} = 0.5$ Hz

Capitolo 3

La missione Voyager

Dopo aver validato per mezzo di segnali simulati il sistema PLL implementato in Simulink, vogliamo capire se lo stesso sistema è in grado di elaborare in maniera soddisfacente segnali reali di missioni deep space.

3.1 Storia della missione Voyager

Nel 1977, dalla base di lancio del Kennedy Space Center a Cape Canaveral, Florida, partirono, a distanza di qualche settimana l'uno dall'altro, due lanciatori Titan III. I due razzi trasportavano le sonde gemelle Voyager 1 e Voyager 2, la cui missione primaria era l'esplorazione di Giove e Saturno. La numerazione indica quale delle due sonde visitò per prima i due giganti gassosi e quale per seconda.

Mentre la traiettoria di Voyager 1, dopo il fly-by di Saturno del 1980, fu piegata dalla gravità del gigante gassoso indirizzando la sonda verso lo spazio interstellare al di sopra del piano dell'eclittica, ossia il piano nel quale giacciono le orbite dei pianeti del Sistema Solare, la Voyager 2, essendo arrivata nel sistema di Saturno circa un anno dopo, nel 1981, sfruttò una fortunata disposizione dei giganti gassosi la cui occorrenza è di circa 175 anni. Tale evento astronomico consentì a Voyager 2 di raggiungere i pianeti Urano e Nettuno per mezzo del *gravity assist*, una tecnica di trasferimento interplanetario che sfrutta l'interazione gravitazionale con un pianeta per piegare la traiettoria di un satellite e fornire a quest'ultimo abbastanza energia cinetica per abbandonare il pianeta, con il vantaggio di risparmiare propellente e tempo di viaggio.

Voyager 2 fu dunque la prima sonda nella storia a visitare tutti e quattro i giganti gassosi del Sistema Solare esterno e l'unica finora ad aver esplorato da vicino Urano e Nettuno. L'incontro con Urano avvenne nel 1986 e dopo tre anni, nel 1989, la sonda raggiunse Nettuno, il pianeta più distante dal Sole. Dopo aver raccolto dati sul pianeta e la sua luna più grande, Tritone, anche Voyager 2 si immise in una traiettoria di sola andata verso lo spazio interstellare, al pari della sua gemella, ma al di sotto del piano dell'eclittica.

Oggi le sonde Voyager sono al di fuori della sfera di influenza gravitazionale del Sole, cioè sono a tutti gli effetti oggetti interstellari, oltre ad essere gli oggetti realizzati dall'uomo più distanti dalla Terra, e continueranno a inviare verso la Terra dati su ciò che si trova al di fuori dell'eliopausa, il confine che separa il nostro Sistema Solare dal mezzo interstellare.

3.2 Architettura e sottosistemi della sonda

In quanto sonde gemelle, la seguente descrizione è valida per Voyager 1 così come per Voyager 2.

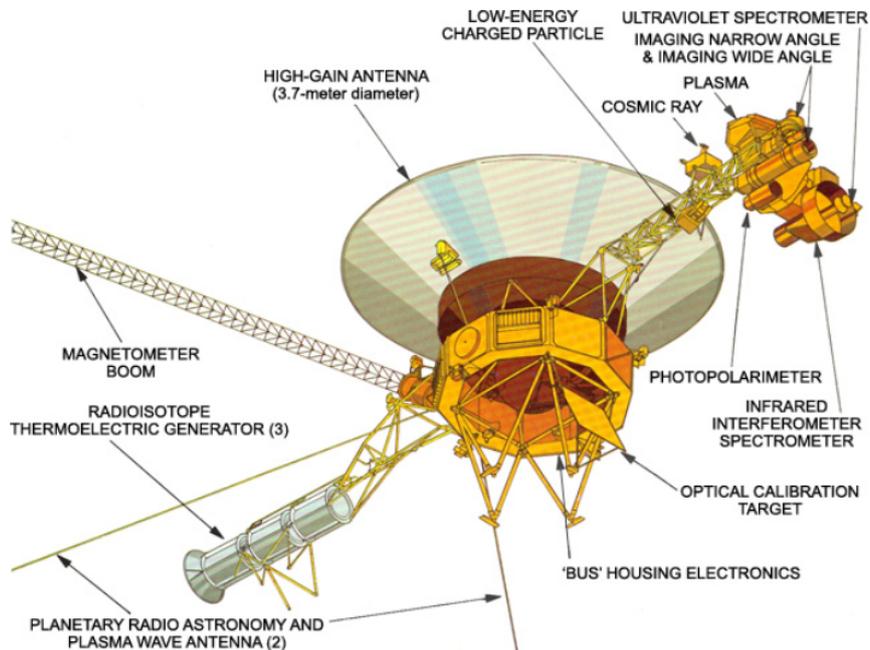


Figura 3.1: Sonda Voyager. Fonte: <https://voyager.jpl.nasa.gov/mission/spacecraft/instruments/>

Il nucleo di tutta la struttura è il bus, il quale ospita i diversi sottosistemi e la strumentazione scientifica. Il bus ha la forma di un grande decagono e l'asse ortogonale al piano nel quale giace tale decagono è anche detto *asse z* e punta sempre verso la Terra, motivo per cui la grande antenna per le telecomunicazioni è allineata lungo tale asse. Sono consentite rotazioni della sonda intorno all'asse *z* attraverso l'accensione di piccoli propulsori collegati al bus. Ognuno dei dieci lati del bus contiene un compartimento che ospita diversi assemblaggi elettronici.

L'antenna ad alto guadagno, o High Gain Antenna HGA, trasmette il downlink su due diversi canali di frequenza, ossia in banda S e in banda X. Mentre il canale in banda X contiene sia dati ingegneristici che scientifici, quello in banda S contiene solo dati ingegneristici sullo stato di salute e di operatività della sonda.

Gli altri sottosistemi comprendono principalmente il magnetometro per le misure di campo magnetico, gli strumenti di telerilevamento, come il Cosmic Ray Subsystem, l'Imaging Science Subsystem, l'Infrared Interferometer Spectrometer and Radiometer, il Photopolarimeter Subsystem e l'Ultraviolet Spectrometer.

La fonte di energia a bordo della sonda Voyager è costituita da tre celle RTG, acronimo di Radioisotope Thermoelectric Generator, collegate in parallelo e montate in tandem su un boom dispiegabile che separa le celle dal resto della struttura per evitare interferenze. Ognuna delle tre celle, infatti, sfrutta il decadimento dell'isotopo radioattivo di Plutonio-238 per generare energia elettrica: l'energia delle particelle alpha rilasciate dal processo di decadimento è convertita in calore, a sua volta convertito in energia elettrica da un convertitore termoelettrico.

3.3 Gli esperimenti nel sistema di Nettuno

L'incontro della sonda Voyager 2 con il sistema di Nettuno ebbe luogo il 25 agosto 1989 e includeva investigazioni di radio scienza sulle masse e le densità di Nettuno e Tritone, le armoniche del campo gravitazionale di Nettuno, la struttura verticale dell'atmosfera e della ionosfera di Nettuno e Tritone, la composizione dell'atmosfera di Nettuno e le caratteristiche dei materiali che costituiscono gli anelli del gigante gassoso. Le osservazioni del sistema di Nettuno comprendono misure di range e misure Doppler per lo studio del campo gravitazionale di Nettuno e della massa di Tritone, misure di occultazione radio dell'atmosfera di Nettuno e Tritone e l'individuazione tramite occultazione radio degli archi d'anello di Nettuno.

In merito all'occultazione, questa consiste nel passaggio di una sonda dietro ad un corpo celeste. La fase in cui la sonda scompare dietro il pianeta è detta *ingress*, quella in cui riappare è detta *egress*. Gli esperimenti di occultazione nel sistema di Nettuno erano basati sulla geometria della traiettoria di Voyager quando la sonda passò dietro Nettuno e Tritone.

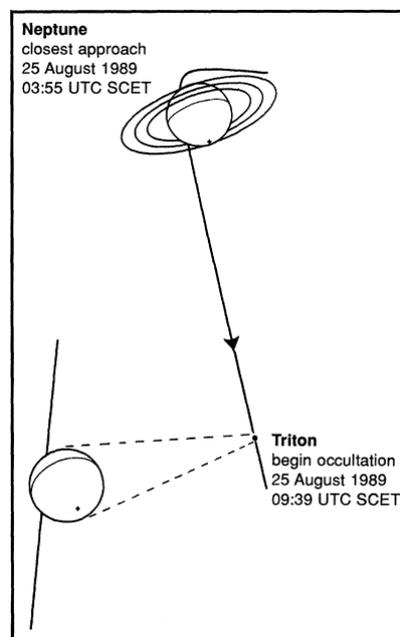


Figura 3.2: Geometria della traiettoria di Voyager 2 per gli esperimenti di occultazione radio nel sistema di Nettuno. L'occultazione di Tritone durò circa 155 s [15]

Per sondare l'atmosfera ed eventuale materiale anulare, il trasmettitore del sottosistema di telecomunicazioni di Voyager inviava segnali nelle bande S e X. Tali segnali erano ricevuti simultaneamente da tre stazioni di tracking: l'antenna di 70 m di diametro della DSS della NASA a Tidbinbilla in Australia, la quale riceveva entrambe le bande di frequenza; l'antenna da 64 m di diametro della CSIRO a Parkes, sempre in Australia, che riceveva solo il segnale in banda X; l'antenna da 64 m di diametro dell'Istituto di Scienze Spaziali e Astronautiche di Usuda, Giappone, che riceveva solo la banda S.

Per quanto riguarda gli esperimenti di occultazione su Tritone, gli effetti atmosferici sono stati individuati e misurati con i dati della fase del segnale sia in ingress che in egress. Dall'analisi dei dati open-loop raccolti dalla stazione di Tidbinbilla, sia l'ampiezza che la fase del segnale esibiscono evidenti effetti di diffrazione in ingress così come in

egress. Per correggere la diffrazione sono state impiegate procedure di filtraggio basate sulla trasformata inversa di Fresnel. Questo trattamento è necessario per diminuire il rischio di confondere le piccole variazioni di fase indotte dalla tenue atmosfera di Tritone con quelle indotte dalla diffrazione. La fase del segnale corretto dalla diffrazione è il primo dato che consente uno studio dell'atmosfera e della ionosfera. Ai fini dell'analisi, le rapide variazioni di fase del segnale ricevuto sono pressoché annullate rimuovendo gli effetti della traiettoria della sonda, dei valori prevedibili della frequenza di riferimento e dei termini di lungo termine, così che le variazioni rimanenti siano solo quelle di origine sconosciuta. Ulteriori variazioni della fase possono essere dovute a:

- variazioni di distanza tra il satellite e la stazione a terra non modellate
- una variazione del contenuto di plasma lungo la direzione satellite-Terra così come del contenuto di gas inerti, quale l'azoto molecolare
- un cambiamento non calibrato della frequenza di riferimento

L'analisi della fase del segnale ha permesso di tracciare dei profili preliminari di densità elettronica e temperatura della ionosfera e di temperatura e pressione dell'atmosfera di Tritone in funzione della quota, ossia della distanza verticale dalla superficie di Tritone [15].

3.4 Analisi del segnale con il PLL

Tramite il modello di PLL implementato in Simulink è stato analizzato il segnale trasmesso da Voyager 2 durante l'esperimento di radio occultazione su Tritone. L'obiettivo dell'analisi è stimare la frequenza media del segnale nel tempo durante le fasi di ingresso e di egress, sia in banda S che in banda X.

Il segnale iniziale ha una frequenza di circa 8.4 GHz per la banda X e di circa 2.3 GHz per la banda S, dopodiché subisce una battitura da RF a IF e da IF alla frequenza residua di circa 1000 Hz, per essere successivamente campionato, prima con una frequenza di 50 kHz e in seguito con un periodo di 3.2×10^{-4} s, vale a dire con una frequenza di 3125 Hz. A causa di una particolarità nella registrazione dei dati per opera del Radio Science Subsystem, il segnale non è in banda base e si verifica una discontinuità nella stima di frequenza da parte del PLL, in particolare questa subisce un "wrap" che la confina tra $-f_N$ e f_N , dove f_N è la frequenza di Nyquist. Occorre pertanto traslare in frequenza lo spettro del segnale così da portarlo in *banda base*. Per ottenere ciò si effettua il battimento del segnale, cioè si sottrae alla sua frequenza una frequenza costante. L'utilizzo dei segnali analitici facilita tale operazione: è sufficiente moltiplicare il segnale analitico che si intende battere con il complesso coniugato del segnale analitico a frequenza costante. Infatti, per le proprietà del prodotto di due potenze aventi stessa base, si ha che:

$$A_{in}e^{j\theta(t)}e^{-2\pi j f_b t} = A_{in}e^{j \int 2\pi f(t) dt}e^{-2\pi j f_b t} = A_{in}e^{j \int 2\pi [f(t)-f_b] dt} \quad (3.1)$$

dove A_{in} è l'ampiezza del segnale analitico in ingresso, $\theta(t)$ è la sua fase, ottenuta per integrazione della frequenza $f(t)$, in accordo con l'Eq. (1.74), mentre f_b è la frequenza costante di battimento. Come mostra l'Eq. (3.1), la funzione integranda all'esponente dell'esponenziale complesso finale è proprio la frequenza del segnale traslata in banda

base di un termine pari alla frequenza di battimento. A questo punto il segnale battuto può essere elaborato senza problemi dal PLL, dopodiché la frequenza media stimata è riportata al suo valore originale trasladola di un termine pari a $f_s - f_b$, essendo f_s la frequenza di campionamento. Nell'ambito di questa trattazione è stata scelta una frequenza di battimento di 50 Hz. La stima di frequenza media è stata effettuata per mezzo del PLL del terzo ordine, essendo questo più performante nel seguire la dinamica del segnale fino a quella parabolica, come mostrato in precedenza. Di seguito sono illustrati i risultati per il segnale in banda X e per quello in banda S. Per i grafici di frequenza è stata adottata la seguente convenzione di colori: in azzurro è indicata la fase di ingress, in arancione quella di egress.

Banda X. Il segnale in banda X è caratterizzato da 5 regioni lungo l'asse temporale:

1. telemetria accesa con CN0 compreso tra 30 e 35 dB-Hz
2. telemetria spenta e ingress di Voyager 2 con CN0 \simeq 45 dB-Hz
3. occultazione della durata di circa 155 s
4. egress di Voyager 2 con CN0 \simeq 45 dB-Hz
5. riaccensione della telemetria con CN0 compreso tra 30 e 35 dB-Hz

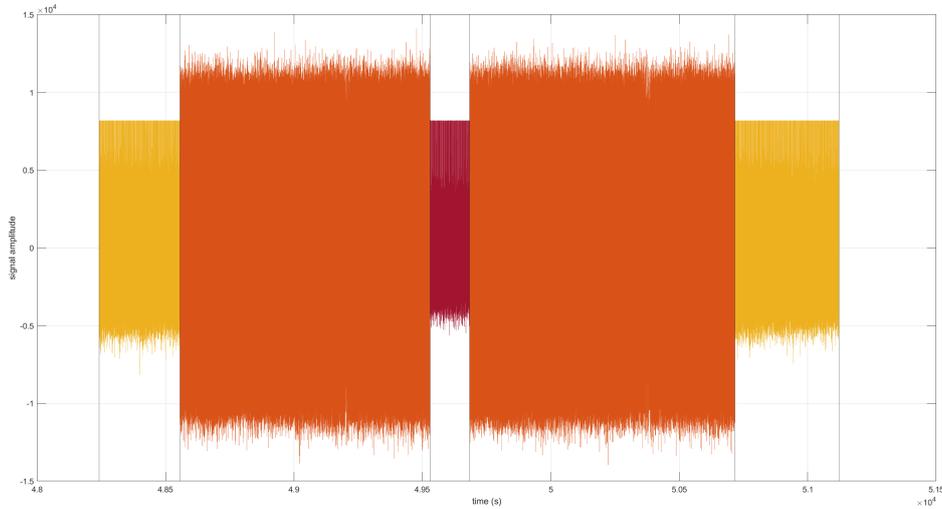


Figura 3.3: Parte in fase *I* del segnale in banda X. Sull'asse delle ascisse è indicato il tempo in secondi, su quello delle ordinate l'ampiezza del segnale. Le regioni dove la telemetria è accesa sono riportate in giallo ocra, le fasi di ingress e di egress sono indicate in arancione e il segnale durante l'occultazione è riportato in scarlatto

La Fig. (3.4) mostra la stima della frequenza residua media nel tempo da parte del PLL, sia in ingress che in egress. Per il calcolo della frequenza media è stato scelto un count time di 1 s e la Loop Bandwidth di ottimo, in questo caso pari a 0.5 Hz.

Infine è stata ricostruita la frequenza ricevuta all'antenna per mezzo della seguente formula:

$$f_{ant,X} = 11f_{POCA} + 7.940 \times 10^9 - 3f_s + f_{rec} + f_{FREQ,X} \quad (3.2)$$

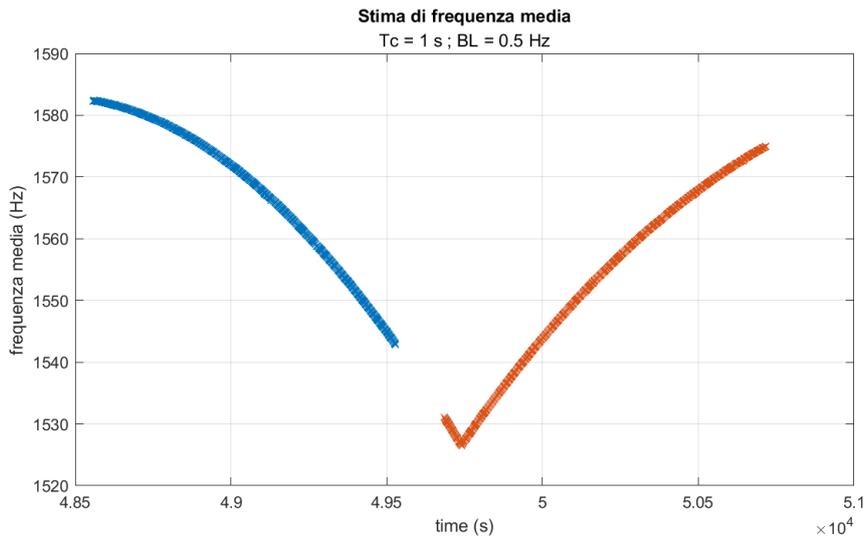


Figura 3.4: Stima della frequenza residua media del segnale in banda X con il PLL del terzo ordine

dove $f_{ant,X}$ è la frequenza in banda X ricevuta dall'antenna, la sigla *POCA* sta per *Programmable Oscillator Control Assembly*, f_s è la frequenza di campionamento di 50 kHz, f_{rec} è la frequenza residua ricostruita tramite il PLL e $f_{FREQ,X}$ è un contributo costante di frequenza.

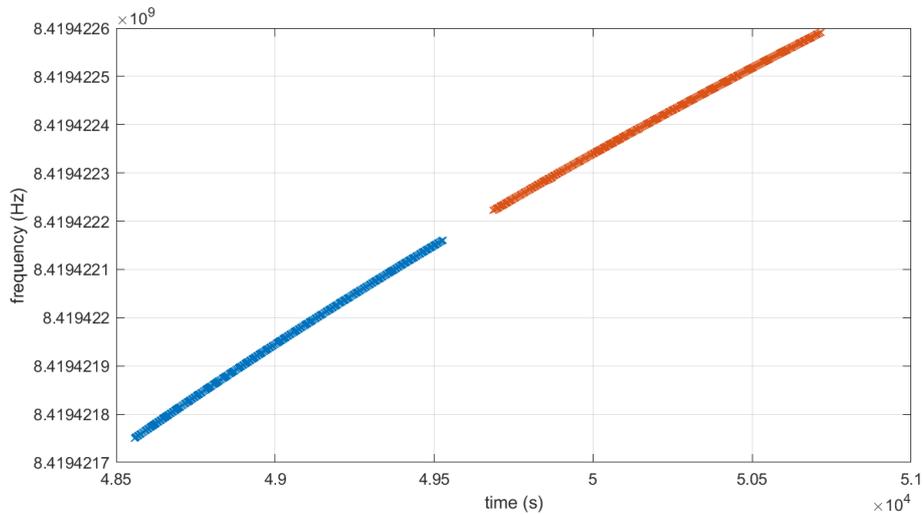


Figura 3.5: Ricostruzione della frequenza ricevuta dall'antenna in banda X

Banda S. Analogamente al caso precedente, il segnale in banda S è suddiviso in 5 parti: telemetria accesa, telemetria spenta e ingress, occultazione, egress, riaccensione della telemetria.

La Fig. (3.6) mostra la stima della frequenza residua media, sia in ingress che in egress. Ancora una volta è stato scelto un count time di 1 s e la Loop Bandwidth di ottimo pari a 0.5 Hz.

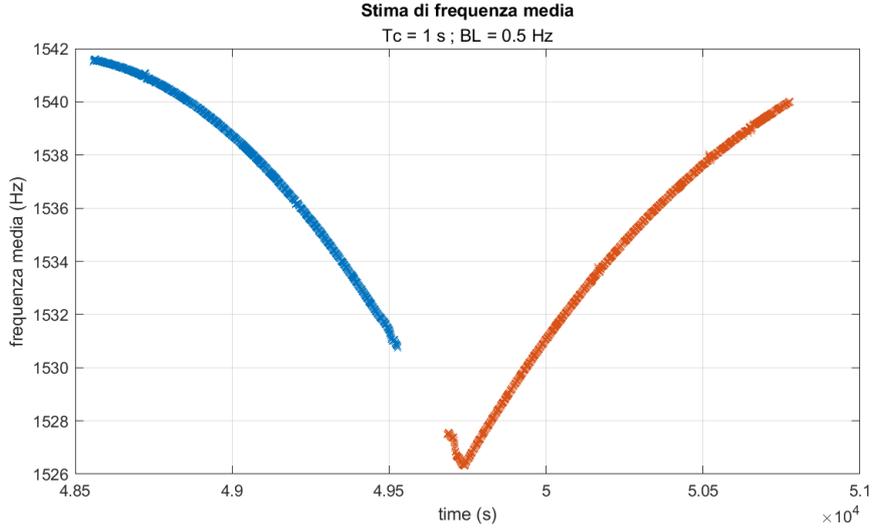


Figura 3.6: Stima della frequenza residua media del segnale in banda S con il PLL del terzo ordine.

In questo caso, la frequenza ricevuta all'antenna è stata ricavata attraverso la formula seguente:

$$f_{ant,S} = 3f_{POCA} + 1.95 \times 10^9 + 3 * \frac{790}{11} \times 10^6 - f_s + f_{rec} + f_{FREQ,S} \quad (3.3)$$

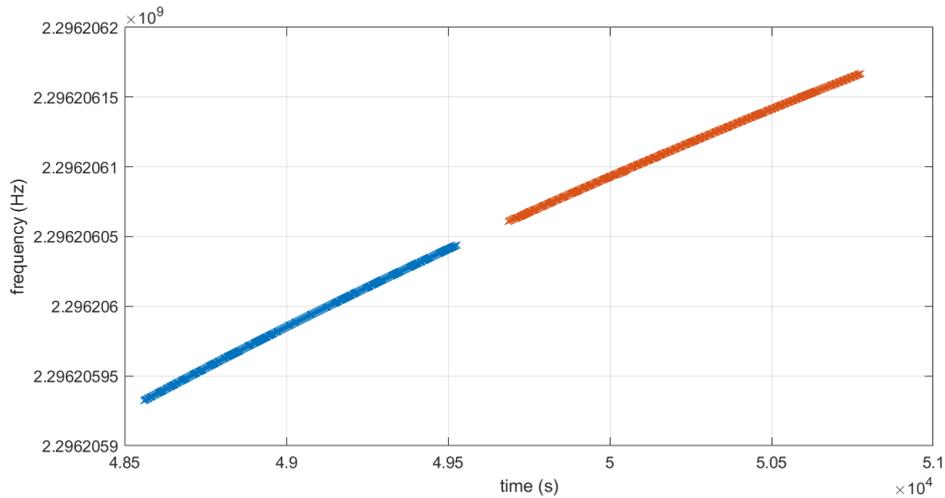


Figura 3.7: Ricostruzione della frequenza ricevuta dall'antenna in banda S

Le stime di frequenza residua media sono state infine confrontate con i risultati ottenuti per mezzo di una tecnica differente basata sulla FFT, ossia l'algoritmo FES [16]. la Fig. (3.8) mostra la differenza tra le stime di frequenza residua media ottenute con il PLL e le stesse stime fornite dall'algoritmo FES.

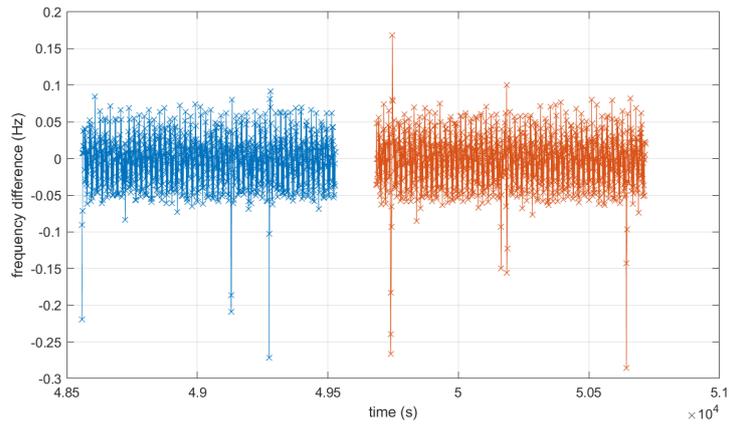


Figura 3.8: Differenza tra la stima di frequenza residua media del PLL e quella ottenuta con l'algoritmo FES

La differenza illustrata in Fig. (3.8) deve essere analizzata non solo da un punto di vista di Digital Signal Processing, ma anche di radio scienza. Al di là della differenza tra le due stime, possiamo affermare che la stima di frequenza media del PLL è buona, considerando che il PLL è fin dall'inizio nella condizione di lock grazie alle condizioni iniziali di fase e frequenza e che la banda di loop utilizzata è quella che teoricamente ottimizza il filtraggio del rumore termico. Quest'ultima considerazione conduce all'ipotesi per cui i picchi che si osservano in Fig. (3.8) non siano da imputare al SNR, ma ad altre cause che devono essere ancora analizzate.

Conclusioni

Al termine di questa attività possiamo affermare che il modello di Phase-Locked Loop implementato in Simulink per digitalizzare le performance di un PLL analogico si è dimostrato efficace, naturalmente entro i suoi limiti di applicabilità, cioè quando riesce a seguire la dinamica del segnale e ad agganciarne la fase una volta assegnate le opportune condizioni iniziali.

Il sistema sviluppato è stato testato con segnali simulati dando esito positivo, dopodiché è stato testato tramite dati reali di una missione deep space con esito altrettanto positivo, anche se rimangono alcune questioni da analizzare. La precisione ottenuta nelle stime di frequenza media è simile a quella ottenuta attraverso i metodi tutt'ora in uso basati sull'analisi spettrale, motivo per cui il sistema implementato sembra essere una valida alternativa per l'analisi dei segnali nell'ambito della radio scienza.

Ringraziamenti

Ringrazio i professori Tortora e Zannoni e il dottorando Togni per avermi seguito durante il lavoro di tirocinio e di tesi.

Ringrazio inoltre i miei genitori per avermi supportato e sopportato fino alla fine.

Un sentito ringraziamento ai colleghi di Ingegneria Aerospaziale e Ingegneria Meccanica per i bei momenti passati assieme e lo spirito di squadra con il quale abbiamo affrontato e superato ogni difficoltà.

Ringrazio infine gli amici, tra cui Barbanti, e gli insegnanti che mi hanno dato la forza di andare avanti.

Bibliografia

- [1] Bob Schutz, Byron Tapley, and George H Born. *Statistical orbit determination*. Elsevier, 2004.
- [2] DW Curkendal and SR McReynolds. A simplified approach for determining the information content of radio tracking data. *Journal of Spacecraft and Rockets*, 6(5):520–525, 1969.
- [3] Christoph Reigber, F Sanso, and R Rummel. Theory of satellite geodesy and gravity field determination. *Lecture notes in Earth sciences*, 25, 1989.
- [4] David A Vallado. *Fundamentals of astrodynamics and applications*, volume 12. Springer Science & Business Media, 2001.
- [5] Catherine L Thornton and James S Border. *Radiometric tracking techniques for deep-space navigation*. John Wiley & Sons, 2003.
- [6] C Chang. Dsn telecommunications link design handbook. *DSN Document*, 373(810–005), 2015.
- [7] Harry Nyquist. Thermal agitation of electric charge in conductors. *Physical review*, 32(1):110, 1928.
- [8] Boualem Boashash. *Time-frequency signal analysis and processing: a comprehensive reference*. Academic press, 2015.
- [9] Steven W Smith et al. The scientist and engineer’s guide to digital signal processing, 1997.
- [10] Floyd M Gardner. *Phaselock techniques*. John Wiley & Sons, 2005.
- [11] Roland E Best. *Phase-locked loops: design, simulation, and applications*. McGraw-Hill Education, 2007.
- [12] AC Densmore. A digitally implemented phase-locked loop detection scheme for analysis of the phase and power stability of a calibration tone. *The Telecommunications and Data Acquisition Report*, pages 207–213, 1988.
- [13] Dustin Buccino, James S Border, William M Folkner, Daniel Kahan, and Sebastien Le Maistre. Low-snr doppler data processing for the insight radio science experiment. *Remote Sensing*, 14(8):1924, 2022.

- [14] Dustin R Buccino, Daniel S Kahan, Oscar Yang, and Kamal Oudrhiri. Extraction of doppler observables from open-loop recordings for the juno radio science investigation. In *2018 United States National Committee of URSI National Radio Science Meeting (USNC-URSI NRSM)*, pages 1–2. IEEE, 2018.
- [15] GL Tyler, DN Sweetnam, JD Anderson, SE Borutzki, JK Campbell, VR Eshleman, DL Gresh, EM Gurrola, DP Hinson, N Kawashima, et al. Voyager radio science observations of neptune and triton. *Science*, 246(4936):1466–1473, 1989.
- [16] Andrea Togni, Marco Zannoni, Luis Gómez Casajús, and Paolo Tortora. An fft-based method for doppler observables estimation in deep space tracking. In *2021 IEEE 8th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, pages 294–299. IEEE, 2021.