Misure Elettroniche

Alberto Tibaldi

8 giugno 2008

Indice

Capitolo 1

Teoria delle Misure

Esistono diverse motivazioni per le quale vale la pena di studiare le misure in modo dettagliato, ossia considerando anche le incertezze e la loro propagazione, basandosi su modelli di vario tipo, che dopo esamineremo. Come mai si sceglie dunque uno studio dettagliato dell'argomento? La risposta è semplice: spesso capita di dover comunicare con altre persone, di doversi scambiare risultati di vario tipo: non è sufficiente avere solo un valore che rappresenti una certa grandezza, perchè serve anche un dato in grado di fornirci l'indeterminazione su questo valore, su questo numero: questa indeterminazione è l'incertezza che tratteremo. Un prodotto per esempio può essere costituito di diversi componenti: poichè esso funzioni, al momento dell'assemblaggio, bisognerà sicuramente aver tenuto conto di quest'indeterminazione, cioè saper determinare di quanto sbaglia ogni singolo componente.

Abbiamo solo finora citato ragioni sostanzialmente ingegneristiche: incertezza come strumento per la quantificazione di un errore, al fine di poterlo comunicare ad un utente, o di poterne tener conto in una misura fornitaci da uno strumento che non conosciamo; esistono ragioni fisiche per le quali la teoria della misura potrebbe essere molto importante: la Fisica è una scienza atta a modellizzare, mediante entità matematiche, il mondo in cui ci troviamo; si tratta dunque di un insieme di modellizzazioni nate da un'osservazione, a partir dalla quale si cerca di esprimere la fenomenologia osservata mediante mezzi matematici più o meno avanzati (questa è quantomeno la base del metodo scientifico fondato da Galileo Galilei). Aver informazioni quantitative sull'errore che commettiamo, permette di affinare il modello sul quale basiamo il mondo, ottenendo informazioni ulteriori, e migliorando ulteriormente la nostra conoscenza della fenomenologia in questione.

In sostanza spesso può capitare di aver a che fare, in prima approssimazione, con un mondo ideale, dove l'indeterminazione non esiste; purtroppo possiamo pensare che ciò non sia vero per diversi motivi:

- Gli strumenti che utilizziamo non sono ideali: qualsiasi strumentazione, digitale o analogica, ha un'indeterminazione intrinseca causata dalla non idealità dei suoi componenti: un condensatore non ideale, un diodo non ideale, piuttosto che qualsiasi altro componente, può provocare errori di misura intrinseci allo strumento più o meno elevati;

- Se anche potessimo considerare ideale il mondo fisico sul quale operiamo, purtroppo esso si discosterà comunque dalla realtà, in quanto la fisica è abituata a lavorare nel vuoto, ossia a pressione nulla; la pressione del mondo reale è tutt'altro che nulla, quindi bisognerà tener conto anche della semplice interazione tra strumentazione e misurando;

- Esiste, molto frequentemente, un certo numero di grandezze di influenza, dotate di incertezza, che per l'appunto modificano lo stato del sistema, attribuendo al sistema misurato stesso un'incertezza;

- Spesso si può avere un problema anche a monte: la stessa definizione del misurando, o del campione di riferimento per la misura, può non essere in grado di ottenere un'indeterminazione nulla.

Il concetto di incertezza e la teoria nascosta dietro ad esso dunque nasce a causa di un insieme di non idealità non eliminabili, neanche con il progredire della tecnologia: per poter garantire un dialogo, una comunicazione tra diversi soggetti, per poter quantomeno arginare i danni portati dal mondo fisico, è stato necessario introdurre queste teorie, e con esse una notazione ben precisa da rispettare, per poter avere garanzie sulla qualità della comunicazione che intendiamo insturare.

1.1 Definizioni

Al fine di poter avere una notazione in grado di permetterci di comprendere il significato di una misura in qualsiasi contesto, è necessario introdurre alcune definizioni fondamentali, per capire quantomeno in modo basilare quali siano i concetti con sui si ha a che fare:

- Sistema misurato: si tratta del sistema sul quale si effettua il processo di misurazione, o di regolazione (se vogliamo progettarlo in modo da soddisfare eventuali nostre necessità); un esempio pratico può essere un amplificatore, del quale vogliamo misurare la frequenza di taglio (come quello di un oscilloscopio, per citare un caso ben noto in Misure Elettroniche);

- Misurando: il parametro che nello specifico intendiamo misurare, nel sistema; esempi sono la frequenza di taglio dell'amplificatore prima citato, o la tensione in una resistenza di un circuito: l'amplificatore e il circuito sono i sistemi, la frequenza di taglio e la tensione i misurandi);
- Misura: si tratta del concetto più delicato e innovativo della teoria che stiamo trattando; la misura associa alle proprietà o caratteristiche degli oggetti una certa quantità numerica; per essere precisi, la misura è formata da tre fondamentali elementi:
  - Valore numerico stimato: il risultato più importante, ossia una stima numerica, ottenuta in un qualche modo (come vedremo più avanti), misurando la grandezza;
  - Incertezza: il possibile discostamento dell'effettivo valore numerico associabile alla grandezza misurata, a partire dal valore numerico stimato: si tratta dunque di una quantificazione dell'errore che possiamo commettere;
  - Unità di misura: ogni grandezza si deve introdurre in un determinato contesto: se pesiamo una massa, o misuriamo la tensione ai capi di una resistenza, effettuiamo sì due misure, ma di grandezze molto diverse; esse si distinguono mediante le unità di misura (che meglio discuteremo in seguito).
- Grandezze di influenza: l'insieme di grandezze che altera la strumentazione utilizzata per la misurazione, e/o l'interazione tra misurando e strumento di misura. Sarà necessario tener conto anche di esse, al momento di formulare la misura, e nella fatispecie la sua incertezza: al variare delle grandezze di influenza, vi è una variazione più o meno sensibile della misura, dunque l'incertezza sulla misura delle grandezze di influenza sarà parte dell'incertezza globale del valore numerico attribuito al misurando. Esempi classici di grandezze di influenza sono la temperatura, o la pressione del sistema.

Sfruttando ciò che abbiamo appena esposto, possiamo riassumere in poche parole cosa significa misurare: il processo di misurazione (o processo cognitivo sperimentale) consiste nell'analizzare una certa fenomenologia, appartenente ad un processo, al fine di poter conoscere il processo stesso. Mediante strumentazioni di vario tipo (teoriche piuttosto che materiali) si discretizza dunque il fenomeno fisico, rendendolo approssimato rispetto a come si presentava ai nostri occhi, ma più semplice da conoscere. A partire dai dati numerici così elaborati (le misure), è possibile studiare il processo, ossia ciò che ha generato le manifestazioni da noi studiate, mediante un'elaborazione dei dati, e quindi un processo di interpretazione dei dati ricavati.

1.2 Classificazione dei metodi di misurazione

Possiamo classificare i metodi di misurazione in diversi modi: una prima classificazione è quella che distingue i metodi in base al numero di letture eseguite, ossia in base al numero di misurazioni effettuate; si parla in questo ambito di:

- Metodi a singola lettura: si effettua una singola lettura su ciascuno degli strumenti coinvolti, in modo da ottenere in uscita un singolo valore, con la sua incertezza e unità di misura; questo tipo di metodo è funzionale soprattutto quando si hanno a disposizione strumenti dotati di scarsa sensibilità: essi non sono infatti in grado di percepire variazioni dovute per esempio a grandezze di influenza, dunque effettuare ulteriori misurazioni sarebbe inutile, in quanto la lettura risulterebbe sempre uguale;

- Metodi a letture ripetute: si eseguono più letture di ogni grandezza in condizioni ipoteticamente uguali, ottenendo come misura il risultato di un'analisi statistica: a partire da un certo numero di letture, ciascuna dotata di incertezza, mediante mezzi statistici si ricava un valore numerico ed un'incertezza.

Vedremo in seguito più dettagliatamente questi metodi di misurazione, specificando meglio le tecniche per la determinazione dei valori numerici e delle incertezze da essi derivanti.

L'altra fondamentale classificazione dei metodi di misura si basa sulla distinzione delle modalità operative di assegnazione della misura ad una grandezza; nella fatispecie, le due possibilità sono:

- Metodi di misurazione diretti: la misura di una determinata grandezza, di un determinato parametro, deriva dall'analisi del parametro stesso: il misurando è effettivamente ciò che intendiamo misurare. Si noti che è possibile dover misurare altre grandezze oltre al misurando, al fine di quantificare correttamente l'indeterminazione: le grandezze di influenza, o comunque qualsiasi grandezza che coinvolge il misurando, va considerata durante il processo di misurazione;

- Metodi di misurazione indiretti: dualmente a prima, il misurando non è il parametro che intendiamo quantificare: ciò che cerchiamo verrà misurato indirettamente, in quanto la sua misura deriverà da un calcolo basato sulle misure di altri parametri, misurati in modo diretto.

Approfondiamo dunque quest'ultima classificazione, lasciando temporaneamente da parte la precedente, che verrà ripresa più avanti parlando di stima dell'incertezza della misura; più avanti verranno inoltre indicati eventuali collegamenti tra le due classificazioni.

1.2.1 Metodi di misurazione diretti

Come già accennato, i metodi di misurazione diretta sono procedimenti di misura che consentono il confronto diretto tra il misurando m ed una grandezza di riferimento, detta anche campione c_i (dove il pedice i indica semplicemente il tipo di metodo di misura che scegliamo, come vedremo, per comodità). Esistono tre diversi tipi di misurazione diretta, che si differenziano come si sarà capito sostanzialmente dal tipo di campione che si sceglie per il confronto diretto con il misurando, come tra poco esporremo. Una breve parentesi riguardante le incertezze dei metodi di misurazione diretta: esistono incertezze che non si possono eliminare, indipendentemente dal metodo di questa categoria che si sceglie di utilizzare: le incertezze legate al misurando, nella fatispecie: quella intrinseca del misurando, ossia l'indeterminazione, la variazione delle caratteristiche del misurando dovute alla misura stessa; quella delle grandezze di influenza, al cui variare modificano lo stato del misurando, aumentando dunque la sua indeterminazione; ulteriore incertezza non eliminabile è quella derivante dall'imperfetta definizione del misurando: non conoscendo esattamente il misurando, è impossibile definirlo con precisione, e quindi con indeterminazione nulla. Un'altra incertezza non eliminabile, anche se leggermente diversa da quelle appena enunciate, riguarda il carico strumentale: l'alterazione del sistema causato dall'interazione campione-misurando. Essa è come già detto non eliminabile, ma varia a seconda del tipo di campione che si sceglie di utilizzare, e quindi a seconda del metodo che si utilizza.

Fatta questa breve introduzione sulle incertezze, analizziamo meglio i tre metodi di misurazione diretta, e le relative fonti di incertezza:

- Misurazione per opposizione: si confronta il misurando m con una grandezza della stessa specie, nota, poichè generata da un campione variabile. A ciò si aggiunge un dispositivo ausiliario, in grado di stabilire una relazione di equivalenza del tipo:

$$m = c_1$$

Le incertezze provocate da un metodo di questo tipo, sono le seguenti: il campione c_1 è dotato di un'incertenza intrinseca, proprio come il misurando m; inoltre, l'equivalenza stessa, $m = c_1$, determina un'incertezza. Un metodo classico di misura per opposizione è la bilancia a due piatti; in elettronica, il Ponte di Wheatstone, oppure i metodi di zero.

- Misurazione per sostituzione: misurare per sostituzione è un metodo basato sul realizzare un campione mediante un campione ottenuto a partire da diversi dispositivi $c_1, c_2, ..., c_n$, che possiamo considerare come parametri di una determinata funzione: il campione c_R è dunque definibile come:

  $$c_R = f(c_1, c_2, ..., c_n)$$

  Si stabilisce a questo punto il rapporto di eguaglianza tra il campione appena ottenuto, c_R, e il misurando m: poichè sia possibile legare le due grandezze, il campione da noi ottenuto mediante la modifica dei vari parametri e il misurando devono essere omogenei. Una volta accertato ciò, possiamo definire la relazione di equivalenza misurando-campione come:

  $$m = c_R$$

  E così si ottiene la misura mediante sostituzione. L'incertezza derivante da questo metodo di misura, ha fondamentalmente due contributi: l'incertezza del campione c_R e la stabilità dei campioni $c_1, ..., c_n$: questi parametri possono essere infatti determinati in modo anche solo approssimativo, a costo che il campione e il misurando siano più omogenei possibile; per realizzare quest'omogeneità, però, bisogna avere parametri stabili, al fine di ridurre le incertezze. Un esempio in elettronica è la sostituzione di una capacità con un'altra, tarata mediante tecniche di risonanza.

- Misurazione con memoria: si tratta del metodo più comunemente utilizzato, specie negli strumenti analogici: il misurando viene misurato semplicemente mediante la lettura da uno strumento (come per esempio una tensione con un voltmetro analogico); ogni strumento è dotato di un particolare diagramma, detto diagramma di taratura: esso rappresenta, per ogni valore di lettura l, il relativo contributo di incertezza dovuto allo strumento (come si legge nei datasheet delle strumentazioni di misura). Parleremo più approfonditamente più avanti della curva di taratura; per ora basti sapere che essa è semplicemente una funzione rappresentante la taratura, ossia una relazione tra la grandezza misurata dallo strumento mediante tecniche di diverso tipo, ed il risultato presentato (con relativa incertezza). Il misurando m viene dunque confrontato con il valore della funzione di taratura, $f_t(l)$, in una certa lettura l_0:

$$m = f_t(l_0)$$

Alle incertezze ricavate dalla curva di taratura, si possono aggiungere errori dovuti all'operatore, nella fatispecie errori di lettura, quali ad esempio gli errori di parallasse.

1.2.2 Metodi di misurazione indiretti

Per metodi di misurazione indiretti, si considerano tutti quei metodi che coinvolgono l'elaborazione di dati, ossia di misure ottenute mediante misurazioni di tipo diretto, ai fini di ottenere una misura. Ottenuto dunque un certo numero di parametri, che elaboreremo mediante relazioni teoriche da noi note a priori, potremo misurare indirettamente un altro parametro a partire da altri che lo coinvolgono in qualche maniera.

Utilizzare un metodo di misurazione indiretto presuppone l'esistenza di un certo modello matematico, che esprima esplicitamente il legame tra i parametri che dovremo misurare direttamente, ed il misurando m. Dovremo dunque misurare n misurandi $m_1, m_2, ..., m_n$, al fine di ottenere, mediante la nostra funzione $f()$, la relazione:

$$m = f(m_1, m_2, ..., m_n)$$

L'incertezza viene stimata a partire dalle incertezze dei singoli misurandi coinvolti nell'operazione di misurazione indiretta: ogni singola incertezza verrà combinata con ogni altra seguendo un processo di elaborazione fondato su alcuni tipi di modelli di propagazione delle incertezze, che analizzeremo più avanti: si parla in questo ambito proprio di incertezza di modello.

Esempio pratico molto semplice e comune di misurazione indiretta è il volume della sfera, a partire dalla sola conoscenza del suo raggio, e di un'incertezza su di esso: mediante una formula geometrica, ed i modelli di propagazione delle incertezze che vedremo tra non molto, sarà possibile ottenere una misura del volume della sfera.

1.3 Stima dell'incertezza di misura

Abbiamo finora proposto un'introduzione alla teoria della misura, enunciando le ragioni per cui esiste, e per quali motivi non è possibile misurare, senza avere un'indeterminazione sul valore misurato. Ai fini di poter soddisfare esigenze di diverso tipo, sono stati presentati due sostanziali modelli matematici a partire dai quali si può calcolare l'incertezza di una misura: il modello deterministico, ed il modello probabilistico. Analizziamoli, e cerchiamo di comprenderne le differenze, e le modalità di uso a seconda della situazione in cui ci troviamo.

1.3.1 Il modello deterministico

Il modello deterministico per la stima dell'incertezza, si può considerare come un modello di worst case: esso rappresenta il metodo più comune per effettuare una sovrastima dell'incertezza, ossia per avere come già si può capire dalle parole worst case, un'incertezza di caso peggiore. Dato un certo valore misurato m_0, mediante il modello deterministico si ottiene una fascia di valore, solitamente simmetrica rispetto al punto m_0, che la misura può avere. Possiamo riassumere in due proprietà il significato del modello deterministico:

- Possiamo ragionevolmente credere che il misurando (o meglio il suo valore numerico reale, ossia il valore numerico che potremmo associare con indeterminazione nulla se disponessimo di un metodo di misura ideale del misurando), rientri nella fascia di valore: dal momento che abbiamo una sovrastima, abbiamo ottime probabilità di avere, tra questi valori, il valore 'vero';
- La fascia di valore è interpretabile come una funzione di distribuzione di probabilità uniforme: tutti gli elementi della fascia di valore hanno eguale probabilità, eguale valore nella rappresentazione del misurando.

Il secondo punto richiede una spiegazione un po' migliore: mediante un modello deterministico otteniamo un valore, ed un range di valori tutti di peso uguale, tutti in grado di rappresentare il valore reale del misurando con la stessa probabilità. Questo è vero anceh quando utilizziamo metodi a letture ripetute: il modello deterministico restituisce sempre e comunque ciò che abbiamo appena esposto con le due proprietà prima enunciate, indipendentemente dalle tecniche di misurazione o di analisi statistica da cui si sono ricavati i dati.

La misura m presentata mediante modello deterministico si presenta nel seguente modo:

$$m = (m_0 \pm \delta)U$$

Dove m_0 è il valore numerico scelto per la misura (normalmente il centro della fascia di valore), δ l'incertezza sul valore, ossia la semiampiezza della fascia di valore, e U l'unità di misura.

Quando si effettuano diverse misurazioni, risulta a questo punto necessario introdurre il concetto di compatibilità: due misure sono compatibili se e solo se le due fasce di valore hanno intersezione non nulla.

Molto spesso capita di dover presentare un'incertezza di tipo diverso: quella che abbiamo sinora esposto è un'incertezza relativa, ossia derivante dal considerare l'errore assoluto, ossia l'errore commesso senza considerarlo a confronto del misurando m_0. Per ragioni di chiarezza, si è deciso di introdurre il valore relativo ϵ di incertezza, ed il valore relativo percentuale ε, utilizzando le seguenti definizioni:

$$\epsilon = \frac{\delta}{m_0}; \varepsilon = \frac{\delta}{m_0} \cdot 100$$

Leggendo l'incertezza relativa abbiamo un'indicazione molto più immediata di quella che ci fornisce l'incertezza assoluta sulla misura: l'incertezza relativa ci dice esattamente di quanto stiamo sbagliando rispetto al valore della misura: se per esempio abbiamo misurato una tensione di 100 V con una $\varepsilon = 2\%$, possiamo subito dire che abbiamo su 100 V un'incertezza di 2 V. Di fatto, parlare di un errore di 2 V non ha molto senso, in quanto potrebbe trattarsi di 2 V su 10 V o di 2 V su 500 V: l'errore assoluto è uguale, ma quello relativo cambia moltissimo.

Esiste un ulteriore modo di esprimere l'incertezza, ossia il metodo mediante valore ridotto: dato un valore convenzionale m_C, come per esempio il fondo scala di un voltmetro impostato ad una certa portata, si può scegliere di proporre l'incertezza nel datasheet in valore ridotto, come:

$$\varepsilon_r = \frac{I}{m_C}$$

Un esempio di incertezza in valore ridotto per esempio è la classe di un dispositivo di misura.

1.3.2 Propagazione dell'incertezza nel modello deterministico

Studiamo come si propagano le incertezze, nel modello probabilistico: avendo diverse fonti di incertezza, vogliamo capire come esse interagiscano tra loro, ai fini di ottenere l'incertezza finale come combinazione di tutte le altre. Distinguiamo a questo punto due sostanziali casi, che il modello deterministico prevede:

- Nel caso di misure ottenute mediante metodi diretti, si esegue semplicemente la somma algebrica dei valori delle incertezze di ciascuno dei contributi introdotti dal metodo di misura (incertezze intrinseche, strumentali, di carico...);
- Nel caso di misure ottenute mediante metodi indiretti, i contributi di incertezza legati alle varie grandezze si combinano mediante le regole che tra poco indicheremo: dato il parametro in misura Y legato dalla funzione f ai diversi parametri misurati $X_1, X_2, ..., X_n$, in questo modo:

  $$Y = f(X_1, X_2, ..., X_n)$$

  Il valore centrale della fascia, y_0, ossia il valore numerico della misura, sarà dato dalla funzione di tutti i valori centrali, $x_{1,0}, x_{2,0}, ..., x_{n,0}$:

  $$y_0 = f(x_{1,0}, x_{2,0}, ..., x_{n,0})$$

  Il valore assoluto dell'incertezza δ_y è dato a partire dalle incertezze assolute $\delta_{x,i}$, ottenute mediante la relazione:

  $$\sum_{i=1}^{N} \left| \frac{\partial f}{\partial x_{i,0}} \right| \delta_{x_i}$$

  Ossia si linearizza mediante sviluppo di Taylor la funzione $f()$ in un intorno del punto $x_{i,0}$ (valore centrale della fascia): derivata della funzione (in modulo, in questo tipo di modello di propagazione), valutata nel punto, moltiplicata per il fattore di peso, il fattore di incertezza.

Si noti che questo tipo di approssimazione è validissimo, a condizione che però i termini di secondo ordine siano trascurabili rispetto a quelli di primo ordine (ossia il fatto che l'approssimazione di f mediante linearizzazione sia una buona approssimazione).

Consideriamo ora la propagazione del nostro modello sulle operazioni principali:

Somma e differenza

Data la misura:

$$Y = X_1 + X_2 - X_3$$

Utilizzando le proprietà appena enunciate, avremo che:

$$\delta_y = \delta_{x_1} + \delta_{x_2} + \delta_{x_3}$$

Ossia: l'incertezza assoluta di una grandezza ottenuta come somma e/o differenza di altre grandezze è pari alla somma delle incertezze assolute delle varie grandezze.

Prodotto e quoziente

Data la misura:

$$Y = \frac{X_1 \cdot X_2}{X_3}$$

L'incertezza assoluta della grandezza Y vale:

$$\delta_y = \frac{x_{2,0}}{x_{3,0}} \cdot \delta_{x_1} + \frac{x_{1,0}}{x_{3,0}} \cdot \delta_{x_2} + \frac{x_{1,0} \cdot x_{2,0}}{x_{3,0}^2} \cdot \delta_{x_3}$$

Consideriamo per ciascuna delle tre componenti un trucchetto: per le prime due moltiplichiamo e dividiamo per la relativa componente, ossia rispettivamente per $x_{1,0}$ e $x_{2,0}$, e nella terza separiamo il denominatore, così:

$$\begin{aligned}\delta_y &= \frac{x_{1,0} \cdot x_{2,0}}{x_{3,0}} \cdot \frac{\delta_{x_1}}{x_{1,0}} + \frac{x_{2,0} \cdot x_{1,0}}{x_{3,0}} \cdot \frac{\delta_{x_2}}{x_{2,0}} + \frac{x_{1,0} \cdot x_{2,0}}{x_{3,0}} \cdot \frac{\delta_{x_3}}{x_{3,0}} \\ &= y_0 \cdot (\epsilon_{x_1} + \epsilon_{x_2} + \epsilon_{x_3})\end{aligned}$$

Ossia, l'incertezza relativa di una grandezza ottenuta come prodotto e/o quoziente di altre grandezze x_i è pari alla somma delle incertezze relative delle altre grandezze x_i.

Potenze, radici

Si possono ricavare queste regole di propagazione semplicemente a partire dalla regola del prodotto.

Data la misura:

$$Y = X^N = \prod_{n=1}^{N} X$$

Applicando la regola del prodotto:

$$\epsilon_y = N \cdot \epsilon_x$$

Per la radice si può far qualcosa di simile; data la misura:

$$Y = \sqrt[M]{X}$$

$$\epsilon_y = \frac{1}{M} \cdot \epsilon_x$$

Piccoli incrementi

Data la misura:

$$Y = X + \Delta X$$

Dove si considera per ipotesi $\Delta X \ll X$

Considerando un ulteriore misurando Z definito come:

$$Z = \frac{Y}{X} = 1 + \frac{\Delta X}{X}$$

Possiamo fare le seguenti osservazioni: utilizzando la definizione di incertezza relativa, possiamo dire che:

$$\epsilon_z = \epsilon_{(1+\Delta x/x)} = \frac{\epsilon_{(1+\Delta x/x)}}{1 + \Delta x/x} \simeq \delta_{(1+\Delta x/x)}$$

Questo poichè consideriamo per ipotesi:

$$\frac{\Delta x}{x} \ll 1$$

Abbiamo quindi l'incertezza di due contributi: 1, e $\frac{\Delta x}{x}$.

Utilizzando la regola della somma, ossia il fatto che data una somma di due grandezze le incertezze assolute van sommate, consideriamo:

$$\delta_{(1+\Delta x/x)} = \delta_1 + \delta_{\frac{\Delta x}{x}} = \delta_{\frac{\Delta x}{x}} = \frac{\Delta x}{x} \cdot \epsilon_{\frac{\Delta x}{x}} = \frac{\Delta x}{x} \cdot (\epsilon_{\Delta x} + \epsilon_x)$$

1.3.3 Il modello probabilistico

Un altro modello per la descrizione della propagazione delle incertezze (in realtà il modello ufficiale in uso, espresso nella GUM (Guide to the expression of Uncertainty in Mesaurement) è il modello probabilistico. Esso è un modello molto più raffinato rispetto al precedente, in quanto fornisce una stima più realistica, con incertezza inferiore, e facendo caso ad alcuni accorgimenti non utilizzabili nel precedente modello.

Potremmo iniziare a chiederci come mai il nome di questo modello è probabilistico: l'enorme differenza rispetto al modello deterministico, sta nel fatto che la grandezza in misura è trattata come una variabile aleatoria, dotata quindi di una propria densità di probabilità. Se prima consideravamo sempre come uniforme la probabilità che un valore piuttosto che un altro fosse il valore reale con il quale si potrebbe rappresentare il misurando, ora si fa un accorgimento di tipo diverso: ciascun valore numerico appartenente alla fascia di valore ha una certa probabilità di essere il valore numerico reale; si introducono dunque concetti di statistica, per introdurre un'espressione della misura anche in questo ambito.

Il valore numerico che si assegna al misurando, in un modello di tipo probabilistico, è rappresentato dal valor medio (o valor atteso, expected value) della variabile aleatoria con la quale modellizziamo la misura.

Esistono a questo punto due fondamentali tipi di incertezze: incertezze tipo, ed effetti sistematici. Consideriamo una misura a valor medio nullo, ossia il cui valore numerico che dovremo presentare nella misura è nullo: la variabile aleatoria rappresentante il nostro misurando è a valor medio nullo. Si definisce come incertezza tipo la deviazione standard dalla media, ossia in questo caso la dispersione dei valori rispetto allo 0. La deviazione standard, come rivedremo, si può intendere semplicemente, in senso statistico, come la radice della varianza della variabile aleatoria. Esiste un altro tipo di incertezze però, come abbiamo appena accennato, ossia gli errori sistematici: essi sono errori che vanno a modificare l'offset della variabile aleatoria, ossia vanno ad influenzare il suo valor medio (come ad esempio il mancato azzeramento di un galvanometro prima della misura). Gli effetti sistematici sono tuttavia eliminabili, se l'operatore è in grado di individuarli e correggerli mediante alcuni accorgimenti nel modello che utilizziamo.

Una nota: nel modello deterministico abbiamo definito il concetto di fascia di valore del misurando come un intervallo all'interno del quale è ragionevolmente compreso il misurando; l'approccio probabilistico di questo modello ci costringe a modificare questo tipo di interpretazione, definendo una nuova fascia di valore, che chiameremo intervallo di fiducia. L'interpretazione di questo nuovo concetto può rappresentare un'estensione del precedente: dato un certo livello di fiducia, ossia una certa probabilità del misurando di trovarsi all'interno di un certo range di valori, questo range di valori è detto intervallo di fiducia.

Mediante questo concetto è possibile, studiando la densità di probabilità della variabile aleatoria che modella il nostro misurando, ottenere un intervallo all'interno del quale, con una certa probabilità, vi sarà il valore numerico reale del misurando. Esprimendo mediante relazioni matematiche, data x la grandezza in misura, e $u(x)$ la sua incertezza tipo, possiamo, studiando la variabile aleatoria che modellizza x, determinare un certo fattore k, detto fattore di copertura, in grado di individuare l'incertezza estesa $U(x)$, in grado di rappresentare l'intervallo di fiducia.

Se consideriamo come valor atteso, ossia come stima del misurando, un punto x_0, e $u(x)$ l'incertezza tipo, allora possiamo dire che l'intervallo di fiducia sarà:

$$[x_0 - k \cdot u(x); x_0 + k \cdot u(x)]$$

Attenzione però: non sempre è possibile ricavare a partire da questi dati il livello di fiducia, ossia la probabilità che effettivamente il valore reale sia all'interno dell'intervallo di fiducia: se non conosciamo la funzione di densità di probabilità modellizzante la nostra grandezza, non siamo in grado di determinarlo.

Esistono come già detto qualitativamente due categorie di incertezze da quantificare, studiando un modello probabilistico:

- Incertezze di categoria A: esse sono incertezze nate da un'analisi di tipo frequentistico e dall'elaborazione statistica dei dati ricavati;
- Incertezze di categoria B: esse sono calcolabili a partire da metodi diversi da quelli frequentistici; nella fatispecie, si possono determinare mediante informazioni note a priori, o forniteci da terze parti.

Approfondiamo l'argomento, cercando di descrivere più nei dettagli questi due tipi di incertezze, ed i metodi per stimarle.

Valutazione di incertezze di categoria A

Supponiamo di indicare con X la grandezza da misurare con un qualche metodo di misurazione, e $x_1, x_2, ..., x_N$ le N osservazioni, ossia i N dati ottenuti in seguito all'uso della prima parte del metodo (quello di raccogliere dati); consideriamo ciascuno dei dati scorrelato dall'altro: ogni osservazione è stocasticamente indipendente dalle altre (parlando in termini probabilistici); possiamo dire che il valor medio (o valor atteso) del nostro fenomeno si possa stimare come:

$$x_0 = \frac{1}{N} \sum_{k=1}^{N} x_k$$

Ossia la sommatoria dei valori, normalizzati per il numero di valori che abbiamo sommato. Come si sa dalla teoria del calcolo delle probabilità, la somma di un grosso numero di variabili aleatorie si distribuisce approssimativamente come una variabile aleatoria normale (gaussiana), secondo il teorema del limite centrale. Inoltre, se intrinsicamente i campioni x_k sono distribuiti secondo una variabile aleatoria normale, allora pur avendo un numero finito di campioni, possiamo dire che la media empirica avrà densità di probabilità gaussiana. Dal momento che abbiamo a che fare con una gaussiana, possiamo ricavare la varianza σ^2, considerando la varianza della nostra distribuzione, standardizzata con la media.

Per poter disporre di un'informazione coerente, scegliamo di fare qualcosa di un po' più particolare, in questo caso, per esporre il risultato che definiremo come incertezza tipo: supponiamo di suddividere i nostri N campioni in $N-1$ insiemi, e di calcolare per ciascuno di essi la media di insieme. La motivazione di questo passaggio è la seguente: il valore numerico che si usa definire nella presentazione della misura effettuata mediante il modello probabilistico, è una media. Vorremmo definire l'incertezza tipo, ossia l'incertezza di questa media, come qualcosa di un po' insolito: dividendo come già detto i nostri N valori in $N-1$ insiemi, e calcolata per ciascun insieme la media di insieme; il valore numerico che in realtà presentiamo nella misura è la media di tutte le medie di insieme (si noti che essa in realtà si discosta pochissimo dalla normalissima media degli N campioni acquisiti durante il processo di acquisizione).

L'incertezza tipo sarà dunque collegata alla varianza di ciascuna media x_i da una fissata media di riferimento x; questa sarà data dall'espressione della varianza dalla media x, $s^2(x)$ che ora indicheremo:

$$s^2(x) = \frac{1}{N-1} \sum_{k=1}^{N} (x_k - x_0)^2$$

La radice positiva della varianza così ricavata, $s(x)$, è detta scarto tipo sperimentale (o più comunemente in ambito statistico, deviazione standard), ed indica il grado di dispersione delle singole osservazioni dalla media di tutte le medie di insieme x_0, ossia dal valore numerico che attribuiremo alla misura:

varianza elevata significa che i campioni sono molto dispersi, e dunque il nostro risultato è poco significativo, in quanto la media difficilmente potrebbe rappresentare il valore numerico effettivo associato al misurando.

Possiamo definire a questo punto la varianza dalla media, a partire dalla varianza empirica, come:

$$s^2(x_0) = \frac{s^2(x)}{N}$$

E, banalmente, da ciò, lo scarto tipo sperimentale dalla media (deviazione standard dalla media) come:

$$s(x_0) = \frac{s(x)}{\sqrt{N}}$$

Abbiamo a questo punto terminato le presentazioni: siamo in grado di definire una misura della grandezza X, quantomeno considerando solo le incertezze tipo A, utilizzando:

- Come valore numerico, la media x_0;
- Come incertezza assoluta tipo (A) u(x), lo scarto tipo sperimentale della media $s(x_0)$;
- L'unità di misura della grandezza in questione.

Ricordiamo però di tutto questo discorso una cosa fondamentale: tutto ciò che abbiamo detto è vero e corretto, a patto che non esista correlazione tra le variabili aleatorie in gioco: tutte le congetture finora effettuate partono proprio da questo principio, e non possono prescindere da esso. Aumentare N potrebbe migliorare in un certo senso la risoluzione, probabilmente diminuire la varianza e quindi l'incertezza, ma anche aumentare la correlazione tra le variabili aleatorie, per questo motivo può essere rischioso. Ciò che abbiamo finora detto può andare a collegarsi con la prima delle classificazioni di metodi di misura, in quanto per effettuare lo studio di incertezze di categoria A si sceglie di raccogliere un certo numero di dati, e di osservare gli effetti che variano infinitesimamente la misura, quali quelli provocati ad esempio dalle grandezze di influenza.

Valutazione di incertezze di categoria B

Abbiamo finora parlato di valutazione di incertezze di categoria A, utilizzando un approccio di tipo frequentistico, e utilizzando mezzi statistici per elaborare dati; la valutazione di incertezze di categoria B non si può assolutamente effettuare mediante strategie di questo genere: esse derivano da accorgimenti da effettuare mediante osservazioni dell'operatore della misura, o a partire da informazioni fornite da terze parti. Un esempio banale di incertezza di categoria B può essere un offset sulla media x_0 provocato da un effetto indesiderato, quale il mancato azzeramento di uno strumento di misura; effetti meno evidenti ma di cui è necessario sempre fare attenzione sono derivanti dalle incertezze dei campioni utilizzati nei metodi diretti di misura: studiando il datasheet di un oggetto che si utilizza per un confronto, o il diagramma di taratura (curva di taratura) di un dispositivo a lettura diretta (da noi anche detto con memoria) si può effettuare una correzione sulle incertezze, considerandone una parte di categoria B.

Non è facile fornire una regola generale in grado di permettere l'individuazione di incertezze di categoria B, poichè esse sono da studiare solo ed esclusivamente nel contesto nel quale ci si trova; tutto ciò che possiamo fare, è tentare di presentare alcuni casi particolarmente comuni, al fine di poterli riconoscere se necessario.

- Data incertezza estesa $U(x)$ e fattore di copertura k noti: è il caso più favorevole: dal momento che l'incertezza di misura dichiarata è stata ottenuta con un fattore di copertura k (come riporta anche il datasheet), allora l'incertezza tipo $u(x)$ si otterrà semplicemente normalizzando per lo stesso k l'incertezza di misura (o incertezza estesa), $U(x)$:

$$u(x) = \frac{U(x)}{k}$$

- Data incertezza estesa e livello di fiducia noti: abbiamo come dato ancora una volta $U(x)$, ed il livello di fiducia. Questo problema è molto più complicato del precedente, in quanto dovremmo a questo punto conoscere la densità di probabilità rappresentante la variabile aleatoria in questione. Avendo ad esempio una gaussiana, possiamo conoscere, dalle tavole degli integrali, i fattori di copertura (per esempio k=1 provoca un livello di fiducia del 67.4 %, k=2 del 95.45 %, k=3 del 99.7 %)

- Data fascia di valore nota: conoscendo la fascia di valore, ad esempio mediante l'indice di classe di uno strumento o alcune formule, è possibile utilizzare il modello deterministico: utilizziamo dunque un caso particolare di modello probabilistico, con una variabile aleatoria uniforme. Considerando l'incertezza assoluta derivante dall'indice di classe dello strumento, δ, sappiamo che l'ampiezza della fascia di valore sarà 2δ, e l'altezza della variabile uniforme, come sappiamo dalla probabilità, $\frac{1}{2\delta}$.

Dalla teoria del calcolo delle probabilità, sappiamo dunque che la varianza di una variabile aleatoria di questo tipo vale:

$$s^2(x) = \frac{\delta^2}{3}$$

E dunque, l'incertezza tipo interpretata come deviazione standard, sarà la sua radice:

$$u(x) = \frac{\delta}{\sqrt{3}}$$

Questo vale per strumenti elettromeccanici di annata, per i quali purtroppo dobbiamo approssimare il modello probabilistico ad un modello deterministico, e utilizzare le sole conoscenze riguardanti la variabile aleatoria uniforme, in questo modo; in strumenti più elaborati, è possibile effettuare uno studio più dettagliato; un esempio può essere per esempio una variabile aleatoria trapezoidale, ossia nel cui centro vi è una probabilità maggiore di trovare il valore numerico effettivo della grandezza misurata. Essa si può pensare semplicemente come un caso particolare della uniforme:

$$u(x) = \delta \cdot \sqrt{\frac{1+\beta^2}{6}}$$

Il parametro β è compreso tra 0 e 1, ed è semplicemente rappresentante la percentuale di base maggiore rispetto alla base minore (una sorta di duty cycle del trapezio: se abbiamo $\beta = 0.5$, abbiamo la base maggiore uguale al doppio della minore; $\beta = 1$ significa avere una densità di probabilità uniforme).

Avendo dunque informazioni sulla fascia di valore (ossia sull'ampiezza dell'incertezza estesa), e sulle caratteristiche della variabile aleatoria (al peggior caso, si utilizza una variabile aleatoria uniforme, utilizzando il modello deterministico), si riesce ad estrarre l'incertezza tipo mediante i procedimenti appena descritti.

1.3.4 Propagazione delle incertezze nel modello probabilistico

Come già fatto per il modello deterministico, consideriamo uno studio della propagazione delle incertezze anche nel contesto del modello probabilistico. Abbiamo visto che è possibile aver a che fare con incertezze tipo A e B: l'incertezza da dichiarare, dopo aver affrontato un processo di misura adottando questo tipo di modello, sarà di fatto una combinazione delle due incertezze (possibilmente attenuando l'effetto delle incertezze tipo B).

Utilizzando un metodo di misura a singola lettura, avremo a che fare esclusivamente con incertezze di tipo B: come già accennato, utilizziamo una singola lettura solo quando abbiamo a che fare con strumenti grossolani, che non sono in grado di percepire variazioni secondarie, quali ad esempio quelle delle grandezze di influenza (con le loro relative fluttuazioni). Possiamo dire che l'incertezza tipo, dato un problema di questo genere (solo incertezze di categoria B), si possa semplicemente calcolare mediante la norma euclidea (radice della somma dei quadrati) di ogni singolo contributo di incertezza tipo B. Possiamo dunque dire che:

$$u(x) = \sqrt{u_{1,B}^2(x) + u_{2,B}^2(x) + ... + u_{N,B}^2(x)}$$

Per quanto riguarda un metodo di misurazione ottenuto a partire da letture ripetute, o mediante metodi indiretti (che coinvolgono relazioni matematiche unenti diverse grandezze misurate), si applicano le regole che ora vedremo: data una relazione Y ottenuta a partire da una funzione di M parametri, di cui H ottenuti mediante letture ripetute, ed i restanti $M - H$ ottenuti mediante metodi a lettura singola (o forniti da terzi):

$$Y = f(X_1, ..., X_H, X_{H+1}, ..., X_M)$$

La stima, ossia il valore numerico introdotto, si ricava mediante due processi diversi (uno per i primi H valori, l'altro per i restanti $M-H$ valori):

1. Per i primi H, si utilizza un metodo statistico: si calcola il valore atteso di ciascuna delle grandezze X_i, ossia $x_{i,0}$;

2. Per i restanti $M - H$ valori, si considera il centro della fascia di valore, $x_{(H+1),C}, x_{(H+2),C}, ..., x_{M,C}$;

La stima y_0 si ottiene come:

$$y_0 = f(x_{1,0}, x_{2,0}, ..., x_{H,0}, x(H+1), C, ..., x_{M,C})$$

Per quanto riguarda invece la stima dell'incertezza di misura, ossia l'incertezza tipo combinata, $u_c(y)$, dobbiamo considerare un'ipotesi notevole, in grado di semplificarci notevolmente i calcoli: l'indipendenza delle misure X_i. Considerando quest'ipotesi di base (come già facemmo precedentemente), possiamo dire che l'incertezza tipo combinata si calcoli come:

$$u_c(y) = \sqrt{\sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i} \right)^2 \cdot u^2(x_i)}$$

Dove $u(x_i)$ è l'incertezza tipo di ogni grandezza X_i misurata, stimata mediante le tecniche di trattamento di incertezza appena viste.

Esiste un modo di calcolare l'incertezza tipo combinata, anche senza l'assenza di correlazione, introducendo un termine correttivo nella radice: è necessario considerare come termine correttivo la covarianza, ossia un termine statistico in grado di quantificare la correlazione, la dipendenza statistica tra variabili aleatorie. Data dunque $u(x_j; x_l)$ la covarianza della coppia di grandezze x_j e x_l; la formula corretta sarà dunque:

$$u_c(y) = \sqrt{\sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i} \right)^2 \cdot u^2(x_i) + 2 \cdot \sum_{j=1}^{M-1} \sum_{l=j+1}^{M} \frac{\partial g}{\partial x_j} \frac{\partial f}{\partial x_l} \cdot u(x_j; x_l)}$$

Presentiamo un effetto interessante, che permetterà di vedere un qualche confronto con il modello deterministico: in statistica si è soliti definire il coefficiente di correlazione lineare $\rho(x_j; y_j)$ come la covarianza normalizzata per il prodotto delle deviazioni standard:

$$\rho(x_j; x_l) = \frac{u(x_j; x_l)}{u(x_j) \cdot u(x_l)}$$

Questa grandezza assume valori compresi tra -1 e +1: se è negativo, significa che vi è una sorta di anticorrelazione: l'aumentare di una grandezza provoca la diminuzione dell'altra; se è positivo, vi è correlazione. Un caso particolare che possiamo analizzare, è il caso di $\rho(x_j; x_l) = +1$: la relazione diventa una cosa del tipo:

$$u_c(y) = \sqrt{\sum_{i=1}^{N} \left[\frac{\partial f}{\partial x_i} \cdot u^2(x_i) \right]^2} = \sum_{i=1}^{N} \frac{\partial f}{\partial x_i} \cdot u(x_i)$$

Questa formula ha un aspetto molto simile a quella presentata per il modello probabilistico, con una differenza: non consideriamo il modulo.

1.3.5 Stima del livello di fiducia

In molte applicazioni, oltre all'incertezza tipo combinata $u_c(y)$, può essere opportuno fornire l'incertezza estesa $U(y)$ calcolata a partire dalla combinata, e da un certo k_p, ossia fattore di copertura k legato al livello di fiducia p.

Non è facile determinare il livello di fiducia corrispondente ad un determinato k_p: possiamo suggerire un metodo probabilistico, molto impreciso: supponendo di aver a che fare con una distribuzione di probabilità normale (come nella maggior parte dei casi effettivamente capita), esistono dei valori di k_p noti (come quelli già citati in precedenza); in caso non si abbia a che fare con una distribuzione normale, però, esiste un'importantissima diseguaglianza, detta diseguaglianza di Chebychev, in grado di fornire un limite superiore alla probabilità che il misurando cada all'interno di un certo range di valori; il problema di questa diseguaglianza è che è molto imprecisa: essa maggiora moltissimo, rispetto al valore effettivo di probabilità, e quindi risulta essere una maggiorazione troppo potente, per quanto sempre verificata.

1.4 Caratteristiche metrologiche di uno strumento di misura

Ai fini di poter effettuare correttamente un processo di misurazione, mediante l'utilizzo di uno strumento di misura, è fondamentale conoscere alcune informazioni riguardo questo. Bisogna dunque disporre di un set di informazioni al fine di poter sfruttare al meglio le prestazioni permesse dal dispositivo, facendo caso alle prescrizioni che bisogna trovare sui manuali. Leggere un manuale non è però sempre semplice: bisogna conoscere la teoria che sta dietro a determinati concetti (che introdurremo tra breve, o abbiamo già introdotto), ai fini di sapere quello che si sta facendo con lo strumento.

In particolare la caratterizzazione metrologica di uno strumento deve essere in grado di fornire legami tra il valore letto sullo strumento, l'incertezza sulla lettura, e le incertezze dovute a elementi esterni quali le grandezze di influenza.

Esistono sostanzialmente due modi, due situazioni in cui comunemente è possibile utilizzare uno strumento di misura:

- Regime statico (o stazionario);
- Regime dinamico.

Consideriamo dunque separatamente questi due tipi di regimi, e le conseguenti definizioni e proprietà di cui bisogna tenere conto.

1.4.1 Regime stazionario

Nel regime stazionario, c'è un certo numero di specifiche, in grado di fornire indeterminazione alla misura; mediante questa introduzione si vuole spiegare in cosa consistono e come analizzarle.

1.4.2 Diagramma di taratura

Il diagramma di taratura di un dispositivo è una relazione in grado di ricavare la fascia di valore da assegnare al misurando, a partire da un certo valore di lettura. In cosa consiste, in parole povere? Come già accennato, effettuando una misurazione con metodo con memoria, la misurazione consiste nella semplice lettura da uno strumento, tarato, calcolando l'incertezza mediante mezzi statistici (e/o a partire dal diagramma di taratura stesso): la curva più evidente, centrale, è la curva di taratura; quelle più esterne riguardano le incertezze dello strumento.

Il diagramma di taratura è semplicemente una relazione in grado di legare il valore percepito da uno strumento di un qualche tipo (elettromeccanico, elettromagnetico, digitale...) ad una scala. La funzione di taratura sulle ascisse presenta proprio questa scala, e sulle ordinate il valore letto dallo strumento, ricavato mediante il suo principio di funzionamento (un tester ICE, strumento elettromeccanico, sfrutta la forza di Lorentz unita alla legge di Hooke su di una molla a spirale, che oppone forza resistiva nei confronti della forza magnetica causata dal moto di cariche). Avere una curva di taratura lineare, significa che ogni trattino della scala è equispaziato al successivo: in qualsiasi punto della scala ci si trovi, la distanza tra i trattini è costante. Con una curva di taratura non lineare, come potrebbe ad esempio essere quella della scala di misura delle resistenze in un tester ICE, avremo i trattini distanziati in modo diverso a seconda della lettura.

Si definisce sensibilità dello strumento sulla lettura l'inverso della pendenza della curva di taratura in un punto (in una lettura): se il diagramma di taratura è lineare, allora la pendenza delle curva di taratura è detta costante di taratura. Se un dispositivo viene presentato e definito mediante la costante di taratura, ossia l'inverso della pendenza della curva di taratura (che sarà una retta), è necessario solitamente presentare come parametro anche la deviazione dalla linearità: essendo molto smeplice trovare effetti di non linearità in dispositivi mediante i quali si son realizzati gli strumenti di misura, discostamenti dalla condizione di linearità potrebbero in effetti risultare piuttosto frequenti. Si definisce campo di misura di uno strumento di misurazione l'intervallo che comprende tutti i valori di misura assegnabili mediante esso; il limite superiore assoluto del campo di misura, viene detto portata dello strumento.

Si parla anche di campo di sicurezza, come estensione del campo di misura: al di fuori di un certo range di valori che lo strumento di misura può percepire (anche se non presentare), è possibile che vi sia una modifica permanente dello strumento a causa di un sovraccarico (ad esempio vi potrebbe essere un sovraccarico sulle molle di un tester ICE e l'isteresi su di esse potrebbe rendere inutilizzabile lo strumento, introducendo troppa corrente). Per questo motivo, andando fuori dall'appena definito campo di sicurezza, si potrebbe rischiare di avere modifiche permanenti del diagramma di taratura del dispositivo di misura.

Ogni diagramma di taratura comprende in sè un certo numero di informazioni, considerando particolari condizioni delle grandezze di influenza (temperatura, pressione, umidità...); si parla di campo di impiego del diagramma di taratura, dal momento che la modellizzazione della curva di taratura, e soprattutto delle relative incertezze, dipende fortemente dallo stato del sistema in cui stiamo misurando, e quindi dalle grandezze di influenza: al di fuori di un certo range di validità, il diagramma di taratura non sarà più in grado di fornirci informazioni adeguate.

Oltre a ciò, bisogna essere in grado di sapere, leggendo le istruzioni o i datasheet dei componenti in uso, che un dispositivo per la misurazione può richiedere un tempo di warm up (pre-riscaldamento), piuttosto che un azzeramento (come nel caso del tester ICE). Facendo caso ad accorgimenti di questo genere, è possibile ridurre le incertezze (soprattutto di categoria B) derivanti da malfunzionamento del dispositivo di misura.

Altre caratteristiche in regime stazionario

Oltre alle già citate sensibilità, portata, ecc., esistono altre caratteristiche, altre proprietà da conoscere parlando di regime stazionario; ne elencheremo qua alcune, discutendole brevemente.

Per risoluzione di uno strumento di misura si intende la minima variazione del misurando in grado di provocare una variazione di lettura pari all'incertezza di lettura. Possiamo di fatto dire che essa sia la minima variazione percepibile dallo strumento e da esso presentabile. In uno strumento analogico dotato di 100 tacchette, la risoluzione è $\frac{1}{100}$ della portata; in uno strumento digitale, è l'ultima cifra significativa sullo schermo. Si noti che risoluzione e accuratezza dello strumento sono concetti molto diversi tra loro: di fatto non è detto che tutte le cifre su di uno schermo di uno strumento numerico siano significative: è possibile certamente leggerle, ma se ad esempio si avesse un'incertezza di 0.1 su di un risultato con 4 cifre decimali, le tre ultime cifre decimali possono essere scartate in quanto non significative. Si può dunque dire che essa sia semplicemente la minima variazione apprezzabile sulla scala di lettura di uno strumento.

L'accuratezza di uno strumento spesso si definisce mediante la classe, ossia l'incertezza espressa in valore ridotto percentuale. Si è soliti considerare come valore di riferimento per la presentazione in valore ridotto la portata, ossia il valore di fondo scala del dispositivo di misura. Si definisce dunque la classe C_L di un dispositivo misurante una grandezza X come:

$$C_L = \frac{\delta X_{FS}}{X_{FS}} \cdot 100$$

Per isteresi di uno strumento si intende la tendenza a fornire valori di lettura diversi in corrispondenza dello stesso misurando, quando esso viene fatto variare per valori crescenti o decrescenti. Modificando dunque il misurando in modo da fargli acquisire valori crescenti o decrescenti, l'isteresi è quella caratteristica del dispositivo di misura in grado di non percepire queste variazioni al di sotto di una certa quantità. L'isteresi, solitamente, varia a seconda del campo di escursione del misurando: supponendo di aver a che fare con un termometro, ad esempio, esso potrebbe avere un'isteresi inferiore su temperature alte, e maggiore per temperature basse.

Per ripetibilità si intende la capacità di uno strumento di fornire letture poco differenti tra di loro a parità del valore del misurando: dato lo stesso misurando, non modificato, nelle identiche condizioni operative, la ripetibilità quantifica gli scarti di misurazione del misurando in diversi momenti.

La stabilità è un concetto simile alla ripetibilità, introdotto in un contesto temporale: si può pensare come una sorta di ripetibilità, considerata in un intervallo di tempo definito: la stabilità è quantificata dall'intervallo di valori di lettura entro il quale si prevede cada una percentuale assegnata delle letture ottenute con lo stesso misurando e nelle stesse condizioni operative, ma in momenti diversi appartententi ad un certo intervallo di tempo. Questo parametro serve a quantificare il deterioramento dell'accuratezza dello strumento col passare del tempo.

1.4.3 Regime dinamico

Le caratteristiche in regime dinamico sono fondamentali sia per lo studio di caratteristiche statiche che di grandezze effettivamente variabili con una determinata ciclicità: quest'ultimo caso è banale, in quanto uno strumento non in grado di percepire determinate variazioni (per esempio segnali a frequenza troppo elevata) non può tornar utile per determinati scopi; in caso di grandezze statiche può però risultare utile caratterizzare il transitorio che precede il regime statico in studio, e così eventuali comportamenti patologici.

In ambito dinamico si è soliti studiare la funzione di trasferimento del dispositivo, $H(j\omega)$, dove j è l'unità immaginaria e ω la pulsazione, anche nota come $2\pi f$.

A partire dalla funzione di trasferimento, si considerano alcune proprietà importanti dello strumento, e una su tutte la banda passante a -3 dB: essa è la frequenza tale per cui si percepisce una riduzione di ampiezza del segnale circa uguale al 30%. Poichè quasi ogni strumento intrinsicamente si può modellare, in analisi in frequenza, come un filtro passa basso, ossia un filtro in grado di tagliare i contributi componenti il segnale al di sopra di una certa frequenza (detta frequenza di cut-off o frequenza di taglio), si sceglie di utilizzare, convenzionalmente, come frequenza di cut-off, la frequenza di taglio a -3 dB.

Altro parametro molto importante per lo studio di un dispositivo di misura è la risposta al transitorio; la funzione di trasferimento intrinsicamente contiene una risposta ad un transitorio di tipo molto particolare, ossia un impulso: antitrasformando la funzione di trasferimento, infatti, si ottiene esattamente una risposta ad un segnale impulsivo. Dalla teoria dei segnali però possiamo pensare ad un transitorio di tipo diverso, come ad esempio un gradino di Heavyside: per questo motivo una delle risposte al transitorio più studiate è la risposta al gradino: analizzando il tempo che impiega il segnale a partire dall'avvio del segnale di gradino (più ideale possibile), si possono studiare funzioni in grado di modellizzare il dispositivo. Nella fatispecie, fondamentale è la formula del transitorio, appresa già in corsi di Elettrotecnica (e motivata in corsi di Teoria dei Segnali):

$$X(t) = (X(0+) - X_\infty)e^{-\frac{t}{\tau}} + X_\infty$$

Ossia, dato $X(0+)$ il segnale, la grandezza al momento di accensione del generatore di gradino, X_∞ il valore a regime del gradino, e τ la costante di tempo (data come $\tau = RC$ con un circuito di primo ordine con un condensatore, o $\tau = \frac{L}{R}$ con un circuito di primo ordine con induttanza), si può così modellizzare una risposta al transitorio, nella fatispecie col gradino.

Capitolo 2

Oscilloscopio Analogico

Incominciamo la nostra trattazione partendo dall'oscilloscopio analogico, spiegando perchè scegliamo di studiarlo, preferendolo a strumenti più moderni ed utilizzati quali l'oscilloscopio digitale; l'oscilloscopio analogico è uno strumento che, per quanto antiquato, rispetto ai moderni oscilloscopi digitali, rappresenta un migliore modello della realtà, e permette di comprendere le idee da cui è nato il concetto, l'idea dietro a questo strumento. Altra motivazione potrebbe essere la seguente: il mondo che osserviamo è analogico, quindi la circuiteria che studieremo sarà analogica, come quella che si può aver studiato in corsi di base quali Elettrotecnica. Senza considerare effetti quantistici, assolutamente non necessari nello studio dell'elettronica, almeno fino a questo livello, il nostro mondo è modellabile in effetti da sole grandezze che variano con continuità nel tempo o nello spazio, quindi studiare strumenti che possano rappresentare questo mondo è senza dubbio meglio, per quest'introduzione al mondo della misurazione, rispetto ad interfacce analogico-digitali che per ora potremmo non essere in grado di comprendere.

L'oscilloscopio analogico è uno strumento in grado di rappresentare in un certo formato un qualsiasi segnale periodico o anche solo ciclico, in qualche modo ripetitivo. Come vedremo, in sostanza è possibile ottenere due tipi di rappresentazioni del segnale: una rappresentazione del segnale al variare del tempo, o anche una rappresentazione delle cosiddette curve di Lissajous, ossia il supporto di una curva presentata come ascissa presa da un segnale $x(t)$ ed ordinata da un segnale $y(t)$ (come vedremo in seguito). L'oscilloscopio permette di effettuare misure, anche quantitative, di studiare forme d'onda, di progettare circuiti e verificarne il funzionamento mediante grafici, di cercare guasti o problemi dei sistemi che progettiamo.

L'immagine dell'oscilloscopio, ossia l'oscillogramma, viene presentata su di uno schermo sul quale è inserita una scala cartesiana, solitamente con 10 divisioni sull'asse delle ascisse, ed 8 sull'asse delle ordinate. Ogni divisione è larga circa 1×1 cm, ed è ulteriormente divisa in 5 parti, in modo da avere una risoluzione di circa 1 mm (considerando anche le semitacchette. Alla base dell'oscilloscopio vi è un tubo a raggi catodici, ossia un lungo tubo a vuoto di vetro sigillato in cui un'estremità emette elettroni che, mediante elettrodi, vengono focalizzati ed accelerati, ed infine direzionati mediante un sistema di deflessione verso lo schermo, dove vedremo di fatto un puntino mobile: esso sarà l'effetto degli elettroni su particolari materiali che si illuminano, quando colpiti dalle cariche in moto.

Incominciamo a discutere dell'effettivo funzionamento di ogni singoli componente dell'oscilloscopio, al fine di presentarne una panoramica completa.

2.0.4 Rappresentazione nel dominio del tempo

Il tubo a raggi catodici si può considerare come il terminale di due canali: il canale verticale ed il canale orizzontale. Essi verranno di fatto trattati separatamente. Notiamo che è presente un commutatore, tra un blocco detto Generatore Base Tempi, ed un ingresso indipendente dal primo canale, per un eventuale segnale esterno, $x(t)$. Questo dipende dal fatto che sono possibili, in un oscilloscopio (a singola traccia), due tipi di rappresentazione: la rappresentazione di un singolo segnale nel dominio del tempo, ossia con il commutatore collegato al Generatore Base Tempi, ed una rappresentazione XY, ossia in cui canale Y e canale X sono alimentati da due diversi generatori, e quindi si ha sullo schermo la rappresentazione parametrica di due segnali (ossia di curve di Lissajous).

Consideriamo ora la rappresentazione di un singolo segnale nel dominio del tempo, cercando di spiegare il funzionamento dello schema a blocchi, motivando tutto ciò che capita. Rappresentazione nel dominio del tempo significa ciò: sull'asse delle x, si ha la progressione temporale, sull'asse delle y il segnale al variare per l'appunto del tempo. Il tempo cresce linearmente, ed il segnale proporzionalmente con la variazione lineare del tempo. Si deve generare, come vedremo, una tensione a rampa, in grado di far scattare la rappresentazione del segnale sullo schermo.

Sull'uscita dell'oscilloscopio y inseriamo il segnale, che sarà di fatto una funzione del tempo. Esso entra in un attenuatore calibrato, ossia in un blocco (vedremo tutto meglio in seguito) che attenua il segnale in modo da adattarlo alla rappresentazione sullo schermo. Esso poi viene amplificato mediante un circuito, e mandato in due punti: da un lato, sulla placchetta di deflessione verticale dell'oscilloscopio, dall'altra al generatore base tempi. Si preferisce pre-amplificare il segnale che finirà sia al sistema di deflessione che alla base tempi, poichè come vedremo in seguito la base tempi è soggetta a molti disturbi, quindi un segnale di una certa intensità può essere molto meglio di uno non amplificato.

Tensione nulla significa colpire il centro dello schermo, tensione minima colpire l'estremo sinistro dello schermo, tensione massima l'estremo destro dello schermo (poichè la tensione di rampa di fatto rappresenta sì il variare nel tempo del segnale, ma quindi anche l'asse delle ascisse del segnale che vedremo nello schermo, e dunque essa sarà, in fin dei conti, semplicemente, la tensione di deflessione orizzontale). Al variare della tensione V_y, si riceve un segnale nel generatore base tempi, che quindi fa scattare la rampa e variare la V_x: così vi sarà deflessione e quindi rappresentazione di una figura sullo schermo. La base tempi, mediante una serie di circuiti, sarà in grado di sincronizzare la rampa con il segnale che invia, e così far rimanere sempre lo stesso segnale sullo schermo. Ricordiamo che per realizzare un oscilloscopio abbiamo detto che servono fosfori con un piccolo tempo di persistenza t_p, ma ciò significa che l'immagine durerà poco tempo sullo schermo. Per avere un'immagine fissa, stabile, dovremo continuare a ripetere la stessa immagine, in modo da far sembrare ai nostri occhi che l'immagine sia di fatto immutabile nel tempo.

La condizione fondamentale per vedere il segnale sullo schermo, è quindi quella che esso si ripeta nel tempo con una frequenza sufficientemente elevata da darci l'idea di essere fisso: il refresh della figura deve essere sufficientemente elevato da non poter essere visto.

2.0.5 Rappresentazione XY

La rappresentazione XY è, sotto il punto di vista dello schema a blocchi, molto più semplice: abbiamo due diversi segnali in ingresso nell'oscilloscopio in due diversi canali, e quindi un segnale al variare del tempo sarà nell'asse y, uno nell'asse x. Il selettore nello schema a blocchi precedentemente inserito sarà dunque posizionato sull'ingresso del canale x, non dipendente da y, ossia senza il generatore di base tempi.

Quella che si vedrà sullo schermo sarà semplicemente il puntino rappresentato dalla coppia (x(t); y(t)). Quella che si ha è quella che in Analisi Matematica si chiama semplicemente parametrizzazione di una curva (curve di Lissajous): al variare del parametro temporale t, tempo relativo al generatore di ciascuno dei due segnali, si ha la rappresentazione sugli assi di diverse curve, tutte parametrizzate rispetto alla stessa variabile:

$$\begin{cases} x(t) = f(t) \\ y(t) = g(t) \end{cases}$$

Dove f(t) e g(t) sono funzioni tra loro indipendenti, ossia ognuna ha il proprio tempo, senza un sistema di riferimento comune. Per poter intuire quale sia l'andamento della curva, bisogna semplicemente cercar di eliminare la parametrizzazione, ossia far in modo di ottenere una funzione del tipo $y = f(x)$ oppure $x = g(y)$. Questo analiticamente non sempre è possibile, ma sono argomenti che non interessano direttamente la disciplina che stiamo trattando.

In questo tipo di rappresentazione ci si propongono problemi di sincronismo: in effetti per quanto abbiamo finora detto il tempo relativo al segnale $x(t)$ ed il tempo relativo al segnale $y(t)$ non sono collegati; ciò significa tuttavia visualizzare sullo schermo punti posizionati in regioni del piano (costituito dallo schermo dell'oscilloscopio) del tutto aleatori; perchè ciò non avvenga, i due generatori devono avere lo stesso parametro temporale, ossia il t di $x(t)$ deve essere lo stesso di $y(t)$; solo a queste condizioni sarà possibile utilizzare le tecniche algebriche ed analitiche, per poter determinare il supporto della curva rappresentante la parametrizzazione dei due segnali.

2.1 Tubo a raggi catodici (CRT)

Il tubo a raggi catodici è il componente che permette di visualizzare il segnale che vogliamo studiare mediante l'oscilloscopio. Esso è l'elemento più importante, se vogliamo, poichè lo schermo sul quale vediamo il segnale, è proprio uno dei terminali del componente (come tra poco meglio vedremo).

Il tubo catodico venne inventato da Crookes nel 1879, e perfezionato da Braun nel 1897, ed è lo stesso componente alla base di tutti i televisori di vecchia generazione (poichè oggi sono completamente obsoleti, in seguito alla nascita della tecnologia LCD); di fatto, per questo motivo, il tubo catodico è il più comune acceleratore di particelle che si possa trovare in una casa. Un tubo catodico è formato da un tubo a vuoto di forma triangolare (un tempo), al cui estremo vi è un catodo (ricordiamo che comunemente nell'elettrochimica per catodo si intende l'elettrodo sul quale avviene un'operazione di riduzione; in una cella elettrolitica esso assorbe elettroni, in una pila ne emette), il quale emette una nuvola elettronica, la quale viene fatta focalizzare in un fascio, e accelerata, fino a raggiungere lo schermo.

Gli elettrodi devono avere un potenziale sempre crescente rispetto al catodo, poichè si deve aumentare notevolmente la velocità degli elettroni emessi dal catodo e convogliati in un fascio dal primo degli elettrodi, al fine di poterli sparare contro lo schermo. Si parla di tensioni nell'ordine dei kV (da 5 a 10 kV circa), e quindi altissime. Per evitare messe a terra indesiderate, al contatto tra utente e schermo, si sceglie di effettuare una piccola astuzia:

anzichè porre lo 0 V sul catodo, si sceglie di metterlo sullo schermo, e porre il catodo ad una tensione fortemente negativa ($\sim -5 \div -10$ kV): così facendo, si eliminano inutili rischi di sicurezza legati all'utilizzo del tubo catodico nell'oscilloscopio.

Descriviamo ora nei dettagli ciò che abbiamo appena sommariamente detto: il catodo è formato da un metallo che, riscaldato, emette elettroni per effetto termoionico. In un intorno del filamento si formerà dunque una nuvola elettronica, la quale potrebbe respingere, con il suo campo elettrico (come secondo la Legge di Gauss), nuovi eventuali elettroni emessi dal catodo. Per questo, la nuvola va immediatamente fatta allontanare dai dintorni del catodo, mediante l'applicazione di una differenza di potenziale positiva rispetto ai -5 kV del catodo, per esempio ponendo una $\Delta V = 500$ V, e quindi inserendo un elettrodo a -4,5 kV, detto Cilindro di Wehnelt, o griglia di controllo. Modificando la tensione della griglia di controllo rispetto a quella del catodo, si decide se accendere o spegnere lo schermo: se la tensione è molto vicina a quella del catodo, la nuvola elettronica resterà accumulata in prossimità dell'elettrodo che non potrà per campo elettrico produrre altri elettroni, e quindi lo schermo resterà spento; se si pone una tensione come, ad esempio, come già detto, -4,5 kV (ossia 500 V in più del catodo), la nuvola elettronica si sposterà e il sistema CRT funzionare. Considerando accesa la griglia di controllo, da qui, si arriverà ad una serie di elettrodi, che dovranno avere una geometria tale da diventare delle sorte di lenti elettroniche: la nuvola dovrà essere convogliata in un singolo punto, al fine di ottenere un singolo fascio che verrà accelerato da altre tensioni su elettrodi successivi, generanti campi magnetici le cui linee di campo fan in modo di raddrizzare sempre meglio il fascio, che verrà sparato verso lo schermo con tensioni man mano crescenti. Arrivati al punto 6, abbiamo un fascio ormai perfetto, che deve però essere direzionato, come una sorta di pennello di elettroni. Per poterli direzionare, ci sono gli elettrodi 7 e 8, elettrodi di deflessione, in grado di deflettere il fascio sull'asse x e sull'asse y, mediante l'applicazione di una tensione ai loro capi. Il vuoto creato all'interno dello schermo permette di aumentare il cammino libero medio dell'elettrone, che così non ha interazioni con altre particelle, e può andare dove noi desideriamo. Per deflettere di 10 cm il fascio è necessaria una tensione di circa 500 V, quindi molto elevata. Questa tensione di deflessione, V_x e V_y, viene applicata non ponendo uno degli elettrodi di deflessione a 0 V e l'altro a V_x (o V_y), ma ponendo sull'elettrodo a tensione inferiore $-\frac{V_x}{2}$ e su quello a tensione superiore $\frac{V_x}{2}$, in modo da non avere andamenti particolari del potenziale che potrebbero esserci scomodi, nella fatispecie dissimmetrie nell'andamento del potenziale.

La sensibilità di deflessione D viene definita come:

$$D = \frac{1}{2}\frac{LV_d b}{dV_{acc}}$$

Dove V_d è la tensione di deflessione, ossia la tensione V_x o V_y, applicata agli elettrodi di deflessione (rispettivamente orizzontale e verticale), V_{acc} è la tensione equivalente all'accelerazione che il fascio ha raggiunto fino agli elettrodi di deflessione, b è la lunghezza del sistema di deflessione, ossia la lunghezza delle placchette, e d la distanza tra le placchette tra di loro. Di fatto il sistema di deflessione è, come possiamo anche notare dall'equazione, molto simile ad un condensatore a facce piane parallele, che però deve avere effetti capacitivi ridotti, al fine di deflettere con efficienza. L'unico fattore sul quale si può dunque agire tra quelli sopra introdotti, al fine di migliorare le prestazioni del sistema di deflessione, è la tensione di accelerazione V_{acc}: facendo entrare nel sistema di deflessione un fascio di elettroni lento, esso subirà maggiormente l'effetto del campo elettrico del sistema, e così potrà essere deflesso maggiormente. Non è invece possibile ridurre la distanza tra le armature poichè il sistema è sostanzialmente un condensatore, e ridurre la distanza tra le sue armature equivale a far aumentare la capacità del sistema, capacità a carico, come vedremo dopo, dell'amplificatore di segnale, provocando effetti indesiderati: ricordiamo infatti che l'introduzione di una capacità riduce la frequenza di taglio del sistema, che diventa così di fatto un filtro passa-basso, cosa che noi vorremmo evitare in un oggetto che dovrebbe rappresentare con la migliore qualità possibile un segnale. Scegliamo dunque di ridurre la tensione di accelerazione al fine di aumentare la deflessione. Si noti però un fatto: ridurre la tensione significa ridurre anche l'energia di impatto sullo schermo, e ciò è negativo: lo schermo è costituito da un sistema elettroluminescente, ossia in grado di rilevare elettroni impattanti sullo schermo con una certa energia cinetica, e trasformarli in segnale luminoso, mediante un sistema di cristalli (che da noi verran sempre volgarmente chiamati fosfori). Perchè però si possa vedere un'immagine sullo schermo, l'energia cinetica degli elettroni deve essere molto elevata: per questo motivo si introduce, dopo gli elettrodi di deflessione, un elettrodo di post-accelerazione: esso produce un campo elettrico radiale, ossia in grado di fornire un'elevata accelerazione, ma senza modificare la direzione del fascio, fornita dal sistema di deflessione. In seguito alla post-accelerazione, gli elettroni giungono sullo schermo, il quale è formato dai fosfori, ossia da questi materiali in grado di illuminarsi se impattati da una particella dotata di notevole energia cinetica, e da una lamina metallica, che avrà due scopi: chiudere il circuito, ossia fare da anodo, da raccoglitore di elettroni, e fungere da specchio, in modo da far vedere sullo schermo l'effetto di luminosità dei fosfori, riflettendo verso l'esterno dello schermo la luce che emettono in seguito all'impatto.

Una caratteristica dei fosfori è quella di avere una persistenza diversa della luminosità, a seconda dei materiali scelti per costruirli: ogni fosforo ha un tempo di persistenza t_p diverso, e a seconda del tipo di strumento che si vuole costruire, ne serviran di diversi. Possiamo dire che il t_p di un radar ad esempio dovrà essere elevato, in modo da poter mantenere a lungo l'immagine degli oggetti sullo schermo; il t_p di un oscilloscopio, in genere, deve essere ridotto. Un test che si effettua spesso sulle caratteristiche di uno strumento, è la risposta al gradino: supponendo di introdurre in uno strumento di misura un gradino, vedere come reagisce lo strumento, ossia come mostra il gradino ideale sullo schermo.

2.2 Generazione base dei tempi

Incominciamo ora a parlare della situazione più complicata da studiare sotto il punto di vista dell'elettronica, nell'oscilloscopio: la rappresentazione nel dominio del tempo. Essa si basa su meccanismi molto più complicati di sincronismo rispetto a quelli che può richiedere una rappresentazione XY (che richiede semplicemente che i due segnali siano sincronizzati sullo stesso tempo); trattiamo ora ognuno dei singoli punti al fine di tentar di fornire una trattazione completa sulla generazione di un segnale nel dominio del tempo.

2.2.1 Trigger

Il segnale derivante dal canale y fornisce al generatore base tempi un impulso in grado di far scattare la generazione di una rampa, ossia di una spezzata rappresentante la progressione del tempo sull'oscilloscopio. Ciò che fa scattare la rampa, ciò che fornisce l'impulso, è il trigger; trigger sta per grilletto, ed effettivamente si può proprio pensare come un grilletto che, quando tirato, fa partire la spazzolata dell'oscilloscopio, permettendoci di vedere sullo schermo la presentazione del segnale. Il circuito di trigger prende comandi da un segnale chiamato sorgente di trigger, che può derivare da tre possibilità:

- INT: segnale prelevato dal segnale stesso che intendiamo osservare, e cioè il segnale di sincronismo viene catturato dal segnale verticale, ed elaborato dal generatore di base tempi, che genera la rampa;
- EXT: si utilizza un segnale esterno all'oscilloscopio, al fine di avere una sorgente esterna di trigger comandata da un altro apparecchio;
- LINE: si utilizza, come segnale di sincronismo, l'alimentazione di rete, e quindi una sinusoide di frequenza 50 Hz.

Il trigger presenta inoltre altri due comandi fondamentali:

- Trigger Level: indica la tensione di partenza, ossia il livello di tensione del segnale che, rilevato dal circuito, deve far scattare l'inizio della rampa;

- Slope: indica la pendenza, positiva o negativa, del punto di scansione: è possibile infatti che un punto allo stesso livello, sia crescente o decrescente, e così si verrebbero a creare ambiguità, e si rischierebbe di non osservare sullo schermo sempre lo stesso segnale. Per questo motivo, si chiede di inserire + se si desidera che la rampa venga lanciata quando il segnale è in una zona crescente, − se si desidera che venga lanciata quando il segnale sta decrescendo.

In pratica, quando settiamo il trigger, scegliamo una tensione di inizio, e scegliamo se desideriamo il segnale in zona crescente o decrescente.

La tensione della rampa generata dal trigger varia tra due limiti: l'estremo sinistro X_{min} e l'estremo destro dello schermo, X_{max}. Sullo schermo di fatto vediamo solamente una rampa per volta; esiste, come meglio diremo dopo, un circuito interno che fornisce il sincronismo, e che, una volta raggiunto, fa partire un segnale di trigger, ossia un segnale che fa scattare l'avvio di una nuova rampa. Questo circuito viene programmato in modo da attendere lo stesso punto del segnale che è stato in precedenza usato per la rampa precedente, dunque può essere inibito al fine di poter partire sempre dallo stesso punto, e quindi disegnare sempre la stessa traccia. In caso contrario, si perderebbe la sincronia tra rampa e segnale, la rampa partirebbe arbitrariamente rispetto al segnale, e sullo schermo non si potrebbe visualizzare il segnale.

Mediante la manopola della regolazione base tempi, è possibile diminuire o aumentare la velocità di scansione, ossia aumetnare o ridurre la pendenza della rampa; se la pendenza della rampa aumenta, sullo schermo verrà disegnata una porzione di segnale inferiore.

La rampa partirà da dal punto X_{min}, e terminerà nel punto X_{max} sull'asse orizzontale, mentre l'asse verticale partirà dal trigger level, seguendo l'andamento del segnale in ingresso nell'oscilloscopio $y(t)$ fino all'esaurimento della rampa, con lo slope scelto, ossia con la pendenza di partenza scelta. La situazione di fine rampa viene avvertita da un circuito interno, in grado di bloccare, rilevato un certo livello di tensione della rampa (X_{max}), il circuito che genera la rampa e resettarlo, portandolo al punto di partenza X_{min}: il circuito generatore contiene infatti un condensatore che viene caricato man mano che la rampa si sposta da sinistra verso destra, e quindi con una sorta di secondo trigger (realizzato mediante un circuito comparatore di soglia) torna al punto di partenza. Si attende allora un nuovo segnale di trigger, che farà ripartire una nuova rampa.

Nella fase intermedia, a partire dal lancio della rampa fino al suo termine (e poco dopo, come spiegheremo), c'è un oscuramento del segnale di trigger e un'inibizione del generatore di rampa, fino a quando non potrà ripartire. Riassumendo, parte la prima rampa, viene inibito il circuito generatore di rampe fino a quando non termina la prima, e quindi, terminato il tempo di blanking (cioè quello di scaricamento del condensatore del circuito generatore), si toglie l'inibizione e si fa ripartire mediante un nuovo segnale di trigger una seconda rampa che verrà visualizzata sullo schermo. Il tempo di inibizione può durare anche più di quello della sola rampa + blanking: in alcune situazioni particolari, può essere necessario impostare un tempo di hold off, ossia un tempo ulteriore di inibizione della rampa, che per particolari casistiche può essere necessario per garantire la sincronia del segnale con le rampe.

Trigger automatico

La procedura di trigger finora introdotta è quella manuale, ossia i cui parametri sono impostati per l'appunto manualmente dall'utente. Talvolta può capitare che però il circuito di trigger non sia in grado di commutare e dunque far partire le rampe, per questo motivo si potrebbe visualizzare lo schermo vuoto. Questo perchè il trigger è un circuito dotato di isteresi: esistono due livelli di commutazione, in grado di far scattare il trigger; uno in salita, uno in discesa. Se avessimo infatti un singolo livello di commutazione, in presenza di rumore, potrebbero partire numerose commutazioni non desiderate, e quindi sicuramente potremmo perdere il sincronismo. Introducendo l'isteresi, e dunque due livelli di commutazione, si può ovviare a questo. Quando però un segnale ha un'ampiezza così piccola da non riuscir a far raggiungere la commutazione al circuito, serve il trigger automatico: le rampe si generano indipendentemente dal segnale di trigger, e dipendono solo dall'hold off e dalla lunghezza delle rampe. Si avrà certezza di avere segnale sullo schermo, ma un rischio è quello di perdere la sincronia del segnale.

Circuito della base tempi

Abbiamo finora parlato in modo molto qualitativo dei circuiti che stanno dietro al funzionamento della base tempi; pur non scendendo nei dettagli troppo profondi dell'elettronica dietro ognuno di questi circuiti, introduciamo quantomeno le idee che stan dietro a tutto ciò che serve per realizzare un generatore di base tempi.

La base tempi si può schematizzare mediante un diagramma a blocchi così:

Il circuito di trigger è formato da un circuito comparatore di soglia, il quale studia continuamente il segnale in attesa di percepire una certa tensione, detta tensione di soglia, e a questo punto permettere al trigger di commutare, e inviare quindi il segnale di generazione. Sul comparatore di soglia agisce la manopola Trigger Level: mediante questa è possibile modificare la tensione di soglia che farà scattare l'impulso. Per poter variare lo slope, vi è un circuito detto circuito invertitore, in grado di invertire il segnale utilizzato come sorgente di trigger, in modo da generare il segnale, se si desidera, con la pendenza negativa.

La rampa sarà generata da una coppia di circuiti: il circuito di flip-flop, ossia un generatore di segnale a gradino sincronizzato con la sorgente di trigger, ed un integratore di Miller, ossia un circuito in grado di integrare il gradino e quindi generare la rampa di ampiezza da noi desiderata. Esso si può comandare mediante una resistenza ed una capacità variabili, per regolare la sensibilità orizzontale, variando la pendenza della rampa. Vi è infine un altro circuito comparatore di soglia, che però agisce in modo diverso, sull'asse delle x: esso è il circuito che avverte la tensione $V = X_{max}$, ossia avverte il fatto che il fascio di elettroni è arrivato al punto più a destra dello schermo, inibisce quindi il circuito e avvia il circuito di blanking, avente il compito di scaricare il condensatore del circuito integratore di Miller. Il circuito di hold-off avrà proprio il compito di inibire, a questo punto, il circuito di flip-flop, al fine di non far scattare ulteriori gradini integrati poi dal circuito di Miller, al fine di evitare problemi di asincronismo.

Nella modalità automatica si utilizza una variante di tutto ciò: il circuito di flip-flop viene utilizzato come ulteriore trigger, in modo da produrre rampe successive al termine del tempo di hold off, ma quindi senza la presenza di segnali sincronizzati con l'ingresso: al termine del tempo di hold-off vengono automaticamente inviati segnali di inizio rampa, tuttavia, se il tempo di hold-off è scorrelato dal periodo di ripetizione del segnale, la rampa parte ogni volta da un punto diverso, sullo schermo dunque si visualizzerà sempre il segnale, ma a partire da un punto differente, e così sullo schermo si vedran immagini sempre diverse, e quindi confusione; per questo motivo, partendo semplicemente dal flip-flop l'impulso di trigger, può non esserci sincronismo, poichè il flip-flop è indipendente dall'ingresso.

Quale può essere una soluzione? Sommare, al segnale di ingresso, un segnale ad esso simile, ossia con le stesse caratteristiche sotto il punto di vista dell'andamento e della frequenza ma un'ampiezza superiore, in modo da permettere la commutazione del trigger, e ottenere quindi un segnale sincronizzato ma anche automatico.

2.2.2 Canale Verticale

Il canale verticale è di fatto rappresentato da tre blocchi in cascata: l'attenuatore calibrato, in cascata con l'amplificatore verticale, in cascata con il sistema di deflessione verticale. L'attenuatore calibrato deve presentare una resistenza caratteristica costante, una bassa capacità, e quindi una banda passante molto elevata. Caratteristiche simili, per quanto riguarda la banda (e gli effetti capacitivi), devono riguardare anche il sistema di amplificazione e di deflessione: la cascata dei tre componenti, come già accennato, è un filtro passa-basso, in quanto taglia le armoniche del segnale al di sopra di una certa frequenza, detta frequenza di cut-off. Maggiore sarà la frequenza di cut-off, migliore sarà la rappresentazione di un segnale sull'oscilloscopio: più infatti sono le armoniche (o frequenze) che passano nel segnale, più la sua rappresentazione sarà simile a quella originale, quella dove non sono stati subiti tagli.

Il fatto che si abbia un effetto capacitivo, comportante un effetto passa-basso, taglie armoniche ad alta frequenza, e così fa perdere velocità nella rappresentazione del segnale, che risulterà essere più lento, o addirittura incompleto, rispetto a quello introdotto nel canale verticale.

Un filtro passa-basso è legato ad un andamento esponenziale di una grandezza: collegando un gradino ideale, sullo schermo dell'oscilloscopio non vedremo il gradino, bensì un esponenziale che lo approssima, a causa degli effetti capacitivi del canale verticale.

Questo si sa dalla teoria dell'elettrotecnica: un condensatore ha un tempo di carica/scarica esponenziale, in quanto l'equazione alla maglia di un circuito RC è di fatto un'equazione differenziale omogenea a coefficienti costanti, che si risolve con una forma del tipo:

$$y(t) = Ae^{-fract\tau}$$

Cerchiamo, a partire da queste considerazioni teoriche, di stimare il tempo di salita del solo oscilloscopio, ossia quello dovuto all'impedenza di ingresso: sappiamo in prima approssimazione che esso è dovuto all'effetto capacitivo del solo oscilloscopio (vedremo che non è esattamente vero):

Vediamo che l'impedenza in ingresso dell'oscilloscopio si può modellare come un resistore R_{in} in parallelo con una capacità, C_{in}. Possiamo pensare che il tempo di salita sia legato dunque alla banda passante del sistema. Sappiamo che la frequenza di taglio del filtro passa-basso interno all'oscilloscopio, f_TO, si può definire analogamente alla banda:

$$f_{TO} = B = \frac{1}{2\pi R_{in}C_{in}}$$

Questa è la banda passante, nonchè la frequenza di taglio del solo oscilloscopio. Si definisce dunque il tempo di salita come il tempo che impiega la tensione per raggiungere il 90% della tensione finale del gradino. Dunque, il tempo di salita intrinseco dell'oscilloscopio, t_{sO}, si definisce così: dato $t_1 : v(t_1) = 0,1V_{max}$, $t_2 : v(t_2) = 0,9V_{max}$,

$$t_{sO} = t_2 - t_1$$

Si può facilmente vedere che, stimando una crescita esponenziale del segnale (come possiamo tranquillamente aspettarci da un sistema passa-basso), il tempo di salita sarà dato da:

$$\ln(0,9V_{max}) - \ln(0,1V_{max}) = \ln\frac{0,9}{0,1} \simeq 2,2$$

A questo punto, poichè:

$$t_{sO} = \frac{1}{f_{TO}}$$

Possiamo dire che:

$$t_{sO} \simeq 2,2R_{in}C_{in}$$

$$f_{TO} = \frac{1}{t_{sO}} = \frac{0,35}{B}$$

Quello che abbiamo appena studiato è un test di un sistema molto utilizzato: la risposta al gradino. Ne avevamo già parlato quando parlavamo dei fosfori, e del fatto che essi hanno un tempo di persistenza dovuto alle proprietà dei materiali di cui son composti; parliamo un po' meglio di questo tipo di test: supponendo di avere un segnale a gradino molto simile e vicino all'idealità, vedere come l'uscita reagirà ad un'eccitazione di questo tipo, è un test eccellente per un tipo di dispositivo/componente. Spesso, al posto del tempo di salita, si considera, in questo ambito, la costante di tempo $\tau = RC$, come indicatore della velocità della funzione a crescere, e dunque della durata del transitorio, della fase esponenziale. Il fatto che si usi la costante di tempo τ ci fa capire che l'oscilloscopio vincoli violentemente il tempo di salita di un segnale: solo segnali contenenti frequenze inferiori a quella di cut-off saranno rappresentati correttamente sul nostro strumento.

Si noti che nella realtà anche il generatore stesso ha effetti capacitivi: un generatore ideale di tensione non ha infatti mai, nella realtà, un'impedenza interna nulla, e quindi è possibile che vi siano effetti capacitivi intrinseci al segnale, quando noi lo colleghiamo all'oscilloscopio.

Il solo gradino potrebbe contenere, ai suoi morsetti, una capacità non nulla tagliante frequenze al di sopra di una frequenza di cut-off, dipendente dalla capacità in questione. Ognuno dei blocchi messi in cascata presenta un proprio tempo di salita, causato da una propria capacità intrinseca. Poichè i dispositivi sono in cascata, dato ad esempio il tempo intrinseco di salita del segnale (causato dagli effetti capacitivi dell'impedenza interna del generatore di segnale), t_{p1}, ed il tempo intrinseco di salita dell'oscilloscopio (causato dagli effetti capacitivi di uno o più di uno dei blocchi in cascata del sistema), t_{p2}, il tempo finale sarà:

$$t^* = \sqrt{t_{p1}^2 + t_{p2}^2}$$

Ossia, quando due blocchi sono in cascata, il loro tempo di salita risultante sarà dato dalla norma 2 dei relativi tempi di salita, ossia dalla loro somma euclidea. Questo vale anche per sotto-blocchi interni: qua abbiamo infatti considerato esclusivamente come blocchi l'intero oscilloscopio e il generatore di segnale, ma se considerassimo l'oscilloscopio come cascata di n elementi, allora la formula diventerebbe:

$$t^* = \sqrt{t_{p1}^2 + t_{p2}^2 + t_{p3}^2 + ... + t_{pn}^2}$$

Il misurista cerca di ridurre innanzitutto l'impedenza di ingresso dell'oscilloscopio, ossia il tempo di salita intrinseco dell'oscilloscopio, t_{p1}, lavorando sul rapporto $\frac{R_{out}}{R_{in}}$; per fare ciò, si utilizza una sonda, ossia un ulteriore blocco in cascata in grado di modificare l'impedenza vista dai morsetti dell'oscilloscopio, e quindi il tempo di salita intrinseco del generatore. Di solito, esistono due categorie di strategie attuabili per lavorare in questa direzione:

- Sonda attiva
- Sonda attenuatrice (attenuatore compensato)

La sonda attiva non è un'ottima soluzione: essa richiede l'uso di un amplificatore ad ampia banda, elemento costoso, e di un'ulteriore alimentazione, anch'essa problematica.

L'attenuatore compensato è una soluzione più utilizzata generalmente in strumenti come oscilloscopi; vediamo come funziona: si tratta in sostanza di un circuito RC che si inserisce tra il generatore di segnale e l'ingresso dell'oscilloscopio; in cascata tra generatore di segnale, ed oscilloscopio, inseriremo dunque un parallelo di un resistore R_S e di una capacità C_S:

L'attenuatore serve a introdurre, all'interno della cascata, un partitore di impedenze:

$$\frac{V_u}{V_i} = \frac{Z_{in}}{Z_S + Z_{in}}$$

La cosa interessante che può introdurre la sonda attenuatrice, è il fatto che, se:

$$C_S R_S = R_{in} C_{in}$$

Allora il partitore di tensione diventa puramente resistivo, si perdono gli effetti capacitivi, e così si ottiene che:

$$\frac{V_u}{V_i} = \frac{R_{in}}{R_S + R_{in}}$$

Questo capita perchè il polo del passa-basso intrinseco dell'oscilloscopio, e lo zero introdotto dall'attenuatore, si compensano, e così si elimina un problema: si ottiene un partitore di tensione resistivo, ma si perde l'effetto passa-basso dell'oscilloscopio, eliminando un tempo di salita, e così in teoria avendo solo più un polo dipendente dal generatore di segnale, sul quale non lavoreremo. Introduciamo dunque la resistenza equivalente R_E e la capacità equivalente C_E di oscilloscopio in ingresso e sonda attenuatrice, che saranno così definite:

$$R_E = R_S + R_{in}$$

$$C_E = C_S // C_{in}$$

Prima di inserire in cascata l'attenuatore appena definito, il tempo di salita, come già detto, valeva:

$$t_s = 2,2 C_{in}(R_{in}//R_G)$$

Con l'attenuatore si avrà un tempo di salita t_s^* così definito:

$$t_s^* = 2,2 C_E(R_{in}//R_E)$$

Qual è il vantaggio in ciò? Generalmente, t_s^* è più piccolo di t_s, poichè $R_E \gg R_{in}$, e dunque, messa in parallelo con la resistenza intrinseca del generatore di segnale, R_O, avremo $R_O//R_E \simeq R_O$. Inoltre, cosa ancora più interessante, $C_E \ll C_{in}$, e possiamo, almeno teoricamente, sperar di aver ridotto il tempo di salita.

Un buon valore di attenuazione del segnale è con una proporzione $10:1$, ossia la sonda attenua di dieci volte il segnale. In questo modo, progetteremo una sonda del tipo:

$$R_S = 9R_{in};\ R_SC_S = R_{in}C_{in} \Longrightarrow C_S = \frac{1}{9}C_{in}$$

In questo modo, sviluppando i conti, troveremo che:

$$t_s^* = \frac{1}{10}T_s$$

Ossia avremo avuto un allargamento della banda pari a 10 volte, e avremo ridotto il tempo di salita di dieci volte rispetto a quello intrinseco dell'oscilloscopio. Osserviamo un fatto interessante: abbiamo finora fatto un test con un segnale di tipo gradino; supponiamo di rifare, con il compensatore, lo stesso test, con un segnale a onda quadra. Ciò ci permette di capire se la nostra attenuazione è stata corretta: infatti, a seconda di come sarà visualizzato il segnale, potremo aver compensato, sottocompensato o sovracompensato il sistema: abbiamo detto che alla base di tutto vi è un filtro passa-basso, che noi, mediante il compensatore, mediante ossia la sonda passiva (sonda compensatrice), appiattiamo introducendo uno zero nel sistema. Se vedessimo un segnale ancora esponenziale, per quanto meno visibilmente, vuol dire che abbiamo sottocompensato il polo: esso è ancora presente nella rete, per quanto meno efficace; duale è la situazione del sovracompensamento: se ora gli impulsi dell'onda quadra fossero decrescenti, avremmo sovracompensato il polo, e cioè lo zero prevarrebbe sul polo, ma provocherebbe l'effetto opposto al precedente: una salita troppo elevata, che tende ad una discesa.

Se l'onda quadra ci appare perfetta, simile all'idealità, allora lo zero della sonda ha compensato alla perfezione il polo, e abbiamo così risolto il nostro problema.

Abbiamo un ulteriore problema: vi è un ritardo, un delay, tra l'istante di trigger e la partenza della rampa. Il trigger infatti manda un impulso di attivazione ad una serie di circuiti, circuiti che comunque hanno un tempo di latenza non nullo prima di avviarsi. Ciò provoca un ritardo nella rappresentazione del fronte, poichè la rampa parte dopo il segnale, e così si perde un pezzo del fronte. In teoria, per ovviare al problema, potremmo cercar di far partire la rampa prima del segnale, ma ciò non ha senso: non può partire la rampa, prima del trigger!

Possiamo agire al contrario: utilizzare un circuito ritardante, in grado di ritardare notevolmente il segnale, in modo da permetterci di visualizzarlo tutto. L'ideale per fare qualcosa di questo tipo sarebbe un circuito dotato di frequenza di taglio tendente a ∞, in grado di fare il nostro lavoro: per qualcosa di questo tipo, si utilizza una linea coassiale.

Queste idee verranno utilizzate per un'ulteriore funzionalità dell'oscilloscopio, che ora descriveremo.

Doppia base tempi

Supponiamo di avere un segnale, del quale vogliamo evidenziare, ingrandire un particolare: la doppia base tempi è stata inventata proprio per permettere una funzionalità di questo tipo. Con la base tempi principale, siamo in grado di visualizzare il segnale normale, intero; con la doppia base tempi, siamo in grado di evidenziare un singolo tratto di segnale, per poi ingrandirlo in un'altra parte dell'oscilloscopio. La prima base tempi sarà dunque la solita, la seconda sarà una base tempi ritardata rispetto alla prima. Anche la seconda base tempi avrà ovviamente bisogno di un circuito di trigger, al fine di poter presentare un segnale mediante una rampa, esattamente come fatto finora; il trigger della seconda base tempi sarà un circuito di delay: mediante un ritardo rispetto alla posizione di partenza della prima base tempi, delay (ritardo) scelto dall'utente, si potrà sincronizzare la seconda base tempi anche a partire dalla prima. In un oscilloscopio, tra le sorgenti della seconda base tempi, si trovan le seguenti opzioni:

- INT
- EXT
- LINE
- BT1

BT1 si riferisce alla prima base tempi: oltre che con le tradizionali opzioni di segnale interno, segnale esterno, linea di alimentazione, è possibile (con un delay aggiuntivo) prendere il segnale di sincronismo dalla prima base tempi, e così partire avendo essa come punto di riferimento. Introducendo un ulteriore Δt, potremo, come già detto, avere il trigger in una posizione a nostra discrezione. Nella visualizzazione, un tratto di segnale avrà un'intensificazione luminosa, e un ingrandimento.

La doppia base tempi ha diverse opzioni di utilizzo, che ora descriveremo brevemente:

- Scansione della sola BT1: sull'oscilloscopio appare soltanto la traccia relativa alla prima base tempi, ossia si considera l'oscilloscopio come se la seconda base tempi non ci fosse, fosse spenta. Si vede l'immagine normalmente, come abbiamo finora studiato, senza alcuna variazione.
- Scansione della sola BT2: sull'oscilloscopio appare esclusivamente la traccia relativa alla seconda base tempi, e quindi si considera solamente la traccia ritardata, solo la traccia della seconda base tempi. Essa si

potrà regolare con le funzioni di delay, o con i soliti controlli, già citati; di fatto anche qua niente di nuovo: l'unica differenza rispetto alle nostre abitudini, è l'utilizzare una base tempi dotata di un opzionale ritardo.

- Scansioni di BT1 + BT2: esistono due opzioni per rappresentare contemporaneamente sia il prodotto della prima che della seconda base tempi; di fatto vedremo sullo schermo il risultato del lavoro di entrambe le base tempi, e non solo più di una, quindi è un tipo di opzione che non abbiamo ancora esaminato finora. Le due opzioni per visualizzare contemporaneamente BT1 e BT2 sono:
  - Modalità MIXED: fino a quando la seconda base tempi non viene attivata, ossia fino a quando siamo in un tempo inferiore al delay, al tempo di partenza della seconda base tempi rispetto al primo, al Δt rispetto all'istante di partenza della prima rampa, sull'oscilloscopio si visualizzerà esclusivamente il prodotto della prima rampa; dal momento che scatterà anche la seconda base tempi, verrà presentata, in sovrapposizione, come una sorta di continuazione dell'immagine, il prodotto della seconda base tempi, regolata in modo da essere ingrandita o rimpicciolita rispetto all'immagine precedente. Si vedrà una cosa del tipo:
  - Modalità Alternate: quella più utilizzata negli oscilloscopi: sullo schermo dell'oscilloscopio si osserva, relativamente alla prima base tempi, il segnale che siamo abituati a vedere, con evidenziata, mediante una sovrailluminazione del tratto interessato, il tratto considerato dalla seconda base tempi. Questo verrà riportato intanto in un altro punto dello schermo, in questo modo: con una spazzolata verrà disegnata la rampa della prima base tempi, con un'altra la rampa della seconda, e, se la frequenza è sufficientemente elevata, i nostri occhi non saranno in grado di percepire le spazzolate e i disegni alternati delle rampe, e visualizzeremo entrambe le basi tempi in questo modo.

Consideriamo un esempio pratico: supponendo di avere un segnale con un disturbo in un determinato punto temporale, dovremo, per poterlo vedere meglio, con un maggior ingrandimento, effettuare i seguenti step:

1. Regolare la prima base tempi in modo da presentare una porzione del segnale che contenga il disturbo;

2. Regolare la seconda base tempi in modo da evidenziare in modo più limitato possibile il disturbo, regolando il delay e la pendenza della rampa.

2.3 Oscilloscopi a doppia traccia

Abbiamo finora parlato di oscilloscopi in grado di presentare un singolo segnale sullo schermo: nella base tempi esso era semplicemente l'ingresso del canale Y sincronizzato sul canale X dal generatore base tempi con le rampe rappresentanti il tempo in crescita lineare, nella modalità XY esso era invece il supporto della curva parametrica le cui coordinate al variare del tempo erano dettate dagli ingressi dei canali X e Y, rispettivamente $x(t)$ e $y(t)$. Non abbiamo ancora pensato ad una possibiità: rappresentare più di un segnale, più di una figura, sullo stesso oscilloscopio. Per poter inserire una funzionalità di questo tipo, si è pensato a due soluzioni:

1. Oscilloscopio a doppio tubo a raggi catodici;
2. Oscilloscopio a commutazione delle due tracce.

La prima soluzione sarebbe valida sotto il punto di vista della qualità, ma gli enormi costi di realizzazione l'hanno fatta accantonare completamente a favore della seconda. Parleremo dunque, in questa sezione, esclusivamente di oscilloscopi a commutazione delle due tracce. A seconda della situazione in cui ci si troverà, si potrà scegliere una delle seguenti opzioni: Alternate, e Chopped.

Modalità Alternate

Nella modalità Alternate, ad ogni spazzolata il commutatore cambierà il canale da cui attingere il segnale da rappresentare: in questo modo, alla prima rampa si rappresenterà il primo segnale, alla seconda il secondo segnale, alla terza il primo segnale, e così via, alternativamente (come suggerisce il nome). Se la frequenza dei segnali è sufficientemente elevata, il nostro occhio non sarà in grado di accorgersi delle spazzolate, e di fatto osserverà i due segnali contemporaneamente sullo schermo, anche se disegnati da spazzolate diverse: ogni rampa disegnerà solo un segnale, e solo al termine di essa la successiva disegnerà l'altro segnale. Per questo, la modalità Alternate va utilizzata solo con segnali ad alta frequenza: segnali a bassa frequenza, ossia lenti, evidenzierebbero il fatto che le rampe per presentarli sono diverse, e quindi si otterrebbe un effetto sgradevole.

Modalità Chopped

Nella modalità Chopped si rappresentano contemporaneamente i segnali sullo schermo, utilizzando come trucco un campionamento asincrono: il fascio elettronico, basandosi su un andamento di tipo onda quadra, con una frequenza intrinseca all'oscilloscopio (e assolutamente indipendente da quella del segnale, di solito fissata intorno ai 100 kHz), salta da un segnale all'altro, rappresentando prima un segnale, poi un altro, in modo da disegnare, con un'unica spazzolata, entrambi i segnali sullo schermo. Il fascio elettronico presenterà esclusivamente i punti dei segnali, quindi non si vedono i passaggi verticali per passare da un segnale all'altro.

Si noti che il segnale dovrebbe dunque apparire sullo schermo tratteggiato, poichè di fatto abbiamo un campionamento, e quindi una discretizzazione dei segnali, e una rappresentazione incompleta, non continua. Ciò non capita in quanto non abbiamo sincronizzazione dei segnali: l'onda quadra disegna, ad ogni spazzolata, tratti sempre diversi del segnale, che così si sovrappongono ai buchi lasciati dalla spazzolata precedente, in modo da darci un'idea di continuità del segnale: la rampa non è sincrona coi campioni, quindi ad ogni spazzolata il punto di partenza è diverso, ed il campionamento avviene su punti diversi; unendo a ciò la lentezza dell'occhio umano, non siamo in grado di notare la discretizzazione dei segnali. Questo tipo di presentazione è dunque ottimale per segnali a bassa frequenza, ossia a frequenza molto inferiore a quella di campionamento: se la frequenza del segnale è confrontabile con quella di campionamento, il campionamento diventa evidente, e così si nota la discretizzazione, e la cattiva presentazione del segnale sull'oscilloscopio.

Problemi di sincronismo

Anche parlando di oscilloscopio a doppia traccia abbiamo a che fare con problemi di sincronismo dei segnali: ora, oltre a tutti i problemi che già abbiam dovuto risolvere, ci si propongono problemi di sincronismo tra le tracce, ma anche di relazione di fase tra esse: se i due segnali sono simili, ossia con la stessa forma d'onda ed isofrequenziali, potremmo infatti voler misurare la relazione di fase tra i due. Come fonti di sincronismo tra i segnali, possiamo scegliere tra tre opzioni:

- INT CH1: considerare come fonte di sincronismo il segnale sul primo canale;
- INT CH2: considerare come fonte di sincronismo il segnale sul secondo canale;
- Commutatore a valle: al termine di ogni rampa, scatta automaticamente una rampa nuova che presenterà l'altro segnale.

Se i segnali sono isofrequenziali, o con una frequenza molto simile, la scelta di sincronismo a monte, ossia considerare uno dei due canali come fonte di sincronismo, è l'ideale: si mantiene tra i due segnali la relazione di fase, in quanto solo un canale farà da trigger, e così il secondo partirà sfasato rispetto al primo, poichè ci sarà lo stesso trigger con lo stesso level e slope per due segnali sfasati intrinsicamente tra loro. Se le due frequenze non sono però identiche (o quasi, quantomeno), un segnale sarà fermo, mentre il secondo continuerà a scorrere sul primo, poichè la non isofrequenzialità porterà ad avere il segnale su cui si basa il trigger come punto di riferimento, e l'altro avrà punti di trigger sempre diversi a causa della differenza di frequenza. Per ovviare a questo problema, esiste la sincronia a valle: se i segnali non sono isofrequenziali, mediante una commutazione a valle, ossia commutazione al termine di ogni rampa, si riesce a presentare entrambi i segnali sullo schermo, ma senza avere una relazione di fase tra i due: non c'è più un segnale che fa da punto di riferimento all'altro, poichè ognuno è il punto di riferimento di sè stesso: si mantiene quindi sempre la sincronia, ma non si può mantenere una relazione di fase tra i due.

Riassumendo, per fare misure di fase è necessario, con segnali isofrequenziali, utilizzare una modalità di trigger basata su uno dei due canali come fonte di sincronismo, ai fini di mantenere la relazione di fase. A seconda della frequenza in questione, converrà utilizzare una modalità di presentazione Chopped (per segnali a bassa frequenza) o Alternate (per segnali ad alta frequenza). In caso di segnali non isofrequenziali, sarà necessario, per presentare entrambi i segnali, usare un sincronismo a valle, con tutti i problemi che ne conseguono.

Capitolo 3

Strumenti Analogici in DC e AC

Quelli che analizzeremo parlando strumenti analogici per la misura di correnti continue o alternate, sono strumenti di natura elettromeccanica: si tratta dunque di strumenti in grado di convertire una tensione, o una corrente, in energia meccanica; mediante una scala basata su questa variazione di grandezze meccaniche, si riesce a quantificare l'energia che l'ha messa in moto, per quanto essa sia di natura non meccanica, e quindi ad effettuare processi di misura.

Lo strumento di base, rappresentante di fatto il cuore di tutti gli apparati che tratteremo, è il galvanometro d'Arsonval; questo strumento è basato sullo studio della forza di Lorentz, agente su di un filo conduttore percorso da corrente I, ed immerso in un campio magnetico $\vec{B}$. Sfruttando i fenomeni legati alla forza magnetica, dunque, si costruisce un circuito basato su di una spira rettangolare, percorsa da corrente, immersa nel campo di cui parlavamo.

Sulla spira si esercita una coppia motrice, causata dalla forza di Lorentz, C_m, quantificabile come:

$$C_m = \vec{F} \cdot \vec{D} = FD\cos(\theta) = IBLD\cos(\theta)$$

Dove F è la forza agente su ciascuno dei lati della spira, formando una coppia, D la lunghezza della spira, L la lunghezza del filo interessato, θ l'angolo che la spira forma con le linee del campo magnetico, considerato uniforme, con tutte le linee di campo radiali: se si vuole rendere costante la coppia C_m, infatti, bisogna rendere $\vec{B}$ parallelo alla spira, in modo da ottenere, per la radialità del campo, $\vec{B}$ sempre ortogonale al piano della spira.

Questo si può ottenere attribuendo alle espansioni polari una simmetria di tipo cilindrico: introducendo all'interno delle espansioni polari con geometria per l'appunto cilindrica un materiale di tipo ferromagnetico, si riesce a migliorare la geometria del campo, direzionando le linee in maniera più radiale rispetto a come non sarebbero naturalmente.

Ciò che abbiamo appena introdotto, è il galvanometro di Arsonval: basandosi su di questo, sarà facilmente possibile costruire strumenti di misura di tensione/corrente AC/DC: introducendo nel nostro circuito-bobina delle molle a spirale, alle quali si collega un indice solidale all'equipaggio mobile (bobina + molle), quando passa corrente nella bobina, l'effetto della bobina è quello di muovere l'indice, quello delle molle di richiamarlo, e così, compensando le due forze, si riesce a quantificare il movimento dell'equipaggio (mediante la risoluzione di una semplice equazione ponente uguale la forza delle molle, alla coppia motrice introdotta dalle bobine).

L'equazione fisica modellizzante la dinamica dell'equipaggiamento mobile, è la seguente:

$$J\frac{\partial^2\theta}{\partial t^2} + K_V\frac{\partial\theta}{\partial t} + K_M\theta = C_M$$

Ossia un oscillatore armonico con equazione completa, il cui coefficiente del termine di secondo ordine è il momento di inerzia dell'equipaggiamento, J, K_V il coefficiente di smorzamento viscoso, K_M il coefficiente di richiamo della molla a spirale utilizzata per lo strumento, e C_m la coppia motrice.

Comportamento Statico

L'equazione appena presentata è la forma più completa in grado di presentare e modellizzare il comportamento del galvanometro; in realtà, è possibile utilizzare forme più semplici, come la seguente, presentante sostanzialmente una caratteristica statica del dispositivo:

$$C_m(I) = B(2NL)IR = BSNI = K_EI$$

$$C_r(\theta) = K_M\theta$$

In condizioni di equilibrio, dunque, i due termini sono uguali, e si ha che:

$$\theta = \frac{K_E}{K_M}I = KI$$

Si noti che questi ultimi risultati contengono la forte ipotesi di considerare solo elementi ideali, e comportamenti ideali e a regime, ignorando fenomeni transitori di qualsiasi genere. Esistono molti, moltissimi elementi di non idealità in un circuito elettromeccanico di questo genere, uno ad esempio costituito dalle molle: esse, oltre a stabilire la posizione dell'organo mobile, devono trasportare una certa quantità di corrente verso di esso. Bisogna essere in grado di regolare però la corrente entrante in queste, poichè una corrente troppo elevata provocherebbe un effetto di surriscaldamento delle molle, e quindi deformazione, distruggendo di fatto il nostro dispositivo.

Comportamento Dinamico

A partire dall'equazione differenziale prima presentata, è possibile ricavare la funzione di trasferimento dell'oggetto che stiamo studiando; essa è una funzione del secondo ordine, e vale:

$$H(s) = \frac{1}{Js^2 + K_V s + K_M}$$

Interpretiamola con un linguaggio più classico; definendo:

$$G = \frac{1}{K_M}; \quad \xi = \frac{K_V}{2\sqrt{K_M J}} \quad \omega_n \sqrt{\frac{K_M}{J}}$$

Possiamo ottenere che:

$$H(s) = \frac{G}{1 + 2\frac{\xi}{\omega_n} s + \frac{s^2}{\omega_n^2}}$$

La forma del transitorio sarà dunque sostanzialmente una cisoide, poichè possiamo intuire che questo sia un transitorio legato ad una funzione di trasferimento con poli complessi coniugati; ξ è il termine di smorzamento del transitorio cisoidale, ω la frequenza di risonanza.

3.1 Strumenti in corrente continua (DC)

Sostanzialmente esistono due tipi di strumenti direttamente realizzabili basandosi sul galvanometro d'Arsonval: l'amperometro, ossia un misuratore di corrente elettrica, ed il voltmetro, ossia un misuratore di tensione, di differenza di potenziale.

3.1.1 Amperometro

Se tarato in maniera idonea, un galvanometro può diventare un misuratore di corrente; innanzitutto, si realizza un galvanometro in grado di misurare correnti molto piccole, nell'ordine dei μA: questi vengono anche detti microamperometri (in realtà non è detto che sian per forza dei microampere, tuttavia ciò come vedremo tra pochissimo permette di disporre di un range di valori misurabili molto più ampio). Questo oggetto avrà una resistenza intrinseca, R_G, modellizzabile come un resistore in serie (e considerando quindi il galvanometro ideale, poichè tutti gli effetti di non idealità, sempre restando in condizioni di regime, sono racchiusi in R_G). Si mette, in parallelo al sistema galvanometro + R_G, una resistenza di shunt, R_S: essa ha lo scopo di modificare la portata dell'amperometro.

Spieghiamoci meglio: il galvanometro, è in grado di reggere e misurare solo una certa corrente, una volta inserito in serie al circuito; questa sarà molto piccola, poichè per ipotesi consideriamo correnti molto basse. Al fine di poter variare la portata dell'oggetto, scegliamo di introdurre, in parallelo ad esso, una resistenza in grado di creare un partitore di corrente tra R_G ed R_S, e così sul ramo del solo amperometro potremo far passare sempre e solo la stessa corrente, ma sull'intero sistema una corrente molto più elevata, che andrà a farsi ripartire tra i due rami. Conoscendo il valore di R_S, e sapendo che R_G è determinato, il partitore è completamente noto, e dunque potremo, conoscendo la corrente che scorre nel galvanometro, sapere quanta corrente è in ingresso nel parallelo. Vediamo infatti che:

$$I_G = \frac{R_S}{R_S + R_G} I \Longrightarrow I = \frac{R_S + R_G}{R_S} I_G$$

Introducendo diverse resistenze, tutte determinate, di valore diverso, è possibile realizzare un amperometro a portata variabile, ossia in grado, mediante un selettore, di cambiare la propria portata. Al fine di realizzare il progetto di un voltmetro, è necessario essere in grado di determinare la R_S idonea alle nostre esigenze; invertendo le formule introdotte:

$$R_S = \frac{R_G I_{G,FS}}{I_{FS} - I_{G,FS}}$$

Dove $I_{G,FS}$ è la massima corrente che può entrare dentro al galvanometro, e I_{FS} la portata equivalente che vogliamo ottenere in ingresso nel parallelo, ossia nell'intero amperometro.

3.1.2 Voltmetro

Per realizzare un voltmetro, ci si comporta in maniera completamente duale: come sappiamo, il voltmetro, al fine di misurare la tensione ai capi di una resistenza, deve essere messo in parallelo ad essa. Come sempre, l'elemento base è un galvanometro, con la sua resistenza R_G in serie; al fine di realizzare un dispositivo con portata maggiore, si introduce un'ulteriore resistenza in serie, R_S, tale per cui si realizza un partitore di tensione tra quest'ultima e la R_G, in modo da ottenere un effetto del tutto identico a quello ottenuto sul voltmetro: variando R_S, si può ottenere una portata maggiore o minore, a seconda delle nostre necessità.

Sappiamo che:

$$V_{FS} = I_{FS}(R_S + R_G)$$

Utilizzando semplicemente la legge di Ohm, e considerando V_{FS} come la tensione di fondo scala, e I_{FS} la massima corrente I_G che può entrare nel galvanometro. Si definisce (impropriamente) un parametro, detto sensibilità del voltmetro, come:

$$K_{\Omega/V} = \frac{R_S + R_G}{V_{FS}} = \frac{1}{I_{FS}}$$

Partendo da qua, si può ricavare semplicemente la formula in grado di determinare la resistenza serie che permette di variare la portata:

$$R_S = K_{\Omega/V} = \frac{V_{FS}}{I_{FS}}$$

Dualmente a prima, mediante selettori di resistenze, è possibile variare la portata dello strumento.

Classe di accuratezza

L'incertezza di questi strumenti, viene di solito dichiarata con un formato ibrido: incertezza relativa ad un valore fisso (che in questo caso è il valore di fondo scala). Si definisce dunque la classe di accuratezza C_L come l'incertezza percentuale relativa al fondo scala:

$$C_L = \frac{\Delta V_{FS}}{V_{FS}} \cdot 100$$

Invertendo la formula, è semplice trovare il valore dell'incertezza assoluta, al fondo scala:

$$\Delta V_{FS} = \frac{C_L V_{FS}}{100}$$

Poichè l'incertezza assoluta, determinata dal parametro ΔV_{FS}, è costante su qualsiasi punto della scala, è possibile facilmente calcolare le incertezze relative al valore letto dallo schermo, mediante le tecniche di stima dell'incertezza:

$$\varepsilon_L = \frac{\Delta V_{FS}}{V_L} \cdot 100 = \frac{C_L \cdot V_{FS}}{V_L}$$

Ciò ci fa intuire che è buona norma leggere valori verso il fondo della portata, ossia verso il fondo scala: se leggessimo valori a inizio scala, l'incertezza relativa sul valore letto, ε_L, aumenterebbe notevolmente (come si può semplicemente evincere leggendo queste formule). Se la lettura è inferiore a $\frac{1}{3}$ del fondo scala, è buona norma cambiare la portata dello strumento in uso.

3.2 Strumenti in corrente alternata (AC)

Mediante alcuni accorgimenti (prevalentemente, di tipo circuitale), è possibile utilizzare tutti gli strumenti sinora analizzati in DC, anche in ambito di correnti alternate, e dunque in AC. Parleremo prevalentemente di voltmetri, ma si sappia sin da subito che tutto ciò che diciamo, è perfettamente esportabile anche in ambito di amperometri, semplicemente cambiando la scala di taratura.

Esistono sostanzialmente tre metodi per effettuare misure in AC, cui sono associati tre tipi di strumentazione (nel nostro caso, tre tipi di voltmetri):

- Voltmetri a valore medio;
- Voltmetri a valore di cresta (o di picco);
- Voltmetri a valore efficace.

Generalmente, come sappiamo da prima, gli strumenti prima analizzati sono solo in grado di lavorare in condizioni statiche, in condizione di DC: se utilizzassimo frequenze vicine a quella di risonanza, ω_n, la funzione di trasferimento acquisterebbe comportamenti imprevedibili dai modelli prima utilizzati, e quindi non si potrebbe più lavorare normalmente con gli strumenti.

Poichè la base degli strumenti analogici resta sempre la stessa, ossia il galvanometro d'Arsonval, sarà necessario cercar di introdurre qualche trucco in grado di ricondurre il nostro problema ad uno studio di tensione DC. Supponiamo di studiare un segnale sinusoidale ideale: nel dominio di Fourier, ossia delle frequenze, esso si può esprimere mediante una delta di Dirac, $\delta(f - f_0)$. A partire da questa, noi vogliamo creare una riga spettrale in grado di rappresentarla alla stessa maniera, ma centrata in 0: considerando la δ centrata in f_0 ampia A_0, vorremmo ricavare la $\delta(f)$, centrata dunque in 0, di ampiezza A_{DC}. Ciò è possibile mediante il seguente artificio: utilizzando un circuito distorcente il segnale, fortemente non lineare, a partire da una singola armonica vengono create armoniche di distorsione di diversa frequenza, tra cui anche una componente continua, ossia un'armonica dislocata a $f = 0$. Introducendo in ingresso un filtro passa basso, tutte le armoniche di ordine superiore alla continua vengono tagliate, e in questo modo, in ingresso al galvanometro, si avrà solo una continua.

Analizziamo ora nel dettaglio ognuno dei tre strumenti finora citati: sostanzialmente essi varieranno, a seconda del filtro, al circuito distorsore, utilizzato per portare la continua al galvanometro, e dunque dal valore di continua scelto per la misura.

3.2.1 Voltmetri a valore medio

I voltmetri a valore medio lavorano nel seguente modo: come filtro, si utilizza un raddrizzatore, ossia uno strumento in grado di tagliare tutta la parte negativa del segnale (ossia, molto banalmente, un diodo in serie ad un resistore): in questa maniera, tutta la parte negativa delle sinusoidi viene tagliata via, e così si ottiene un segnale positivo. La continua contenuta in questo segnale raddrizzato, rappresenta il valor medio utilizzato dal voltmetro: un filtro passa basso taglierà la componente alternata, e farà dunque solo passare la componente rappresentante l'offset introdotto dalle sinusoidi raddrizzate.

Data la tensione $v(t)$ definita come:

$$v(t) = V_P \sin\left(\frac{2\pi}{T}t\right)$$

Il valor medio, V_m, sarà:

$$V_m = \frac{1}{T}\int_0^T |v(t)|\, dt$$

Questa formula, in realtà, fa riferimento a qualcosa di diverso: qua consideriamo il modulo del segnale, $|v(t)|$, poichè consideriamo non solo di tagliare la parte negativa, ma anche di raddrizzarla.

Esistono infatti due possibilità (di cui noi ne abbiam citata una sola, ossia il diodo in serie al circuito), di raddrizzamento di una forma d'onda (sinusoidale):

- Raddrizzamento a singola semionda: esso consiste, come già detto, nella semplice introduzione di un diodo in serie al circuito, in grado di tagliare tutta la parte negativa. In questo caso, il valore $V_{m,1}$ fornito al galvanometro, sarà:

$$V_{m,1} = \frac{1}{T} \int_0^{\frac{T}{2}} \left(V_p \sin \left(\frac{2\pi}{T} t \right) dt \right) = \frac{V_p}{\pi}$$

- Nel caso di raddrizzamento a doppia semionda, si effettua l'equivalente della funzione valore assoluto: la parte positiva del segnale viene ribaltata, ossia ne viene considerato il simmetrico rispetto all'asse delle x. in questo modo, il vlaor medio $V_{m,2}$, fornito al galvanometro, sarà:

$$V_{m,2} = \frac{2}{T} \int_0^{\frac{T}{2}} \left(V_p \sin \left(\frac{2\pi}{T} t \right) dt \right) = \frac{2V_p}{\pi}$$

Si noti un dettaglio: quello che noi abbiamo così ottenuto, è il valor medio della tensione; per questioni di convenzione, di solito, il costruttore tara lo strumento in termini di valore efficace, considerando il fatto che:

$$V_{eff} = \frac{V_m}{\sqrt{2}}$$

In questo caso, semplicemente:

- Nel caso di strumenti con raddrizzamento a singola semionda, la scala dello strumento sarà presentata moltiplicata per un fattore circa uguale a $2,22$:

$$V_{eff,1} = V_{m,1} \cdot 2,22$$

- Nel caso di strumenti con raddrizzamento a doppia semionda, la scala dello strumento sarà presentata moltiplicata per un fattore circa uguale a $1,11$:

$$V_{eff,2} = V_{m,2} \cdot 1,11$$

Si noti che questo tipo di presentazione è stato ideato solo ed esclusivamente per grandezze di tipo sinusoidali, per segnali di tipo sinusoidale. Si noti che dunque, se avessimo segnali di tipo non sinusoidale, questo tipo di presentazione del valore non sarebbe significativa: lo strumento, a prescindere dal tipo di segnale in ingresso, continua a moltiplicare sempre per lo stesso coefficiente (che sarà o 2,22 o 1,11 a seconda di singolo o doppio raddrizzamento); con segnali non sinusoidali, prima di tutto è necessario dividere il valore letto per il coefficiente utilizzato dallo strumento, dopodichè, determinata una relazione tra valor medio e valor efficace propria del segnale in questione, si può effettuare la conversione.

3.2.2 Voltmetri a valore di cresta

Circuitalmente parlando, il filtro non lineare rappresentante il distorsore per ricavare l'armonica continua da inviare al galvanometro, non è molto differente dal circuito usato per ricavare il valor medio, anche se l'effetto finale è molto diverso.

Come elemento raddrizzatore utilizziamo un diodo, ma in parallelo all'ingresso introduciamo un condensatore. Lo scopo di questo circuito, come vedremo tra poco, è quello di trasformare il segnale in ingresso, sinusoidale, nel suo valore di cresta, ossia nel suo valore massimo: la continua che si estrae dunque dal segnale non è più il valor medio, bensì il valor massimo della sinusoide.

Spieghiamo come funziona il circuito, considerando gli elementi ideali: all'accensione, per $t = 0$, il condensatore è scarico, e dunque il diodo è polarizzato direttamente, quindi conduce. La tensione ai capi del condensatore, in questo primo momento, segue quella ai capi del condensatore, che risulta essere in fase di carica. Una volta carico, al livello massimo V_M, l'anodo del diodo tende a perdere tensione, ma poichè il condensatore (considerato ideale) è ormai carico, mantiene la tensione a V_M anche per quel poco tempo in cui il diodo ha tension inferiore a quella di accensione, mantenendo dunque costante la corrente.

Nella realtà, il condensatore non è ideale, ed il voltmetro neanche (ha cioè resistenza finita): se la resistenza del voltmetro non è infinita, una piccola parte della corrente esce dal condensatore, e così, quando il diodo diviene interdetto, passa un minimo di carica nel ramo del voltmetro, poichè quello del diodo è off. Riuscendo tuttavia a realizzare un circuito molto poco reattivo, ossia con una τ molto elevata, si riesce comunque a ottenere risultati decenti.

Si noti che il voltmetro non è l'unica fonte di non idealità del circuito: il diodo di fatto non è ideale, e ciò comporta l'introduzione di un'ulteriore perdita, poichè, al momento dell'interdizione, passa corrente di segno oposto, e quindi elimina parte della carica presente.

Esiste una strategia alternativa: a seconda del fatto che la qualità del diodo sia o meno superiore a quella del condensatore, si può scegliere di invertire le loro posizioni nel circuito rilevante il valore di cresta, in modo da rilevare la tensione, anzichè sul condensatore, sul diodo.

Si noti un grosso problema di questo circuito, dovuto alla sua topologia ed ai suoi componenti: se il segnale sinusoidale non fosse a media nulla, ma presentasse un offset, questo non verrebbe in alcun modo rilevato: il condensatore tende a tagliare la continua, e quindi il valore che rileveremmo non sarebbe quello della tensione sinusoidale e dell'offset assieme, ma solo il valor massimo della sinusoidale: di fatto quello che fa il circuito è traslare la sinusoide ponendo il suo massimo a $v(t) = 0$, e lavorare di qui; a prescindere dagli offset, il valor massimo non subirà variazione. Questo tipo di circuito è detto di clamp: il segnale pone la cresta a 0, annullando gli effetti della continua, e ricavando quindi solo le informazioni del segnale sinusoidale.

La scala verrà tarata, nel caso di segnali sinusoidali, normalizzando come al solito per $\sqrt{2}$, come prima avevam discusso. Poichè i voltmetri presentano sempre il valore efficace, riferito però a segnali di tipo sinusoidale, sarà necessario moltiplicare la misura per $\frac{\sqrt{2}}{2}$, in modo da ritrovare il V_M, e quindi determinare manualmente il valore efficace.

3.2.3 Voltmetri a valore efficace

Finora, abbiamo calcolato in vari modi una continua, basandoci su sistemi di diverso tipo, e a partir dal valore ricavato abbiamo calcolato, mediante alcune leggi, il valore efficace V_{eff}. I voltmetri a valore efficace sono più interessanti (nonchè più usati degli altri, molto spesso), in quanto misurano, a prescindere dalla forma d'onda del segnale, il suo valore efficace, V_{eff}. Il valore efficace è una definizione basata a partire dalla definizione di potenza media, che dunque andrà ripresa:

$$P = \frac{\frac{1}{T}\int_T |v(t)|^2 \, dt}{R}$$

Legato a questa definizione di potenza, si introduce dunque la definizione corretta di valore efficace come:

$$V_{eff} = \sqrt{\frac{1}{T}\int_T |v(t)|^2 \, dt}$$

Come già detto, una delle cose veramente interessanti di questo strumento, è il fatto che è utilizzabili in forme d'onda di qualsiasi tipo: sinusoidali, non sinusoidali, distorte, rumorose (ossia rumori di diverso tipo). Se i due tipi di strumenti prima presentati non sono in grado di presentare per come sono realizzati in maniera corretta il risultato, strumenti di questo genere presentano come risultato il valore efficace, e lo presentano correttamente qualsiasi sia il segnale.

Esistono due categorie di voltmetri a valore efficace: una basata su principi elettromeccanici, in grado di realizzare strumenti robusti, perfetti per applicazioni industriali, ma assolutamente non accurati; la categoria che utilizzeremo noi, da laboratorio, per studiare elettronica di segnale, si basa su elaborazioni analogiche del valore efficace V_{eff}, mediante circuiti abbastanza elaborati, di diverso tipo.

Questo tipo di sviluppo analogico di V_{eff} si basa sostanzialmente sulla definizione di valore efficace, implentata mediante blocchi in grado di effettuare il quadrato del segnale, di integrarlo, e di far la radice dell'integrale appena ricavato.

Altro schema piuttosto interessante è la conversione elettrotermica, che ora vedremo più nel dettaglio.

Conversione elettrotermica

Dati due resistori identici, anzichè misurare direttamente la tensione efficace, si calcola l'equivalente tensione continua in grado di mandare il resistore alla temperatura a cui si trova, a causa del passaggio di corrente. Si tratta della soluzione più utilizzata in ambito di voltmetri a valore efficace.

Entrando meglio nel merito, sappiamo che una tensione $v(t)$, ai capi di un resistore R, provoca una variazione (incremento) di temperatura $\Delta\theta$, per effetto Joule. Questa tensione è alternata, ma noi potremmo valutare comunque l'incremento di temperatura, considerando, anzichè l'alternata, la continua equivalente in grado di effettuare questa variazione di potenza dissipata dal resistore in calore. Possiamo dire che:

$$\frac{V_{eff}^2}{R} = \frac{V_{DC}^2}{R}$$

E quindi:

$$V_{eff} = V_{DC}$$

Lo schema di principio di questo tipo di voltmetro si basa su di un amplificatore differenziale.

Riscaldando un resistore, si avrà un riscaldamento di una delle giunzioni dell'amplificatore differenziale, o meglio delle giunzioni dei BJT utilizzati per realizzare l'amplificazione; si viene a creare uno squilibrio dettato da una variazione di temperatura di uno dei due transistor, e quindi vi sarà una variazione delle tensioni tra base ed emettitore. Questo segnale è una tensione continua, e sarà diversa da quella dall'altro capo dell'amplificatore, poichè solo da una parte introdurremo una variazione termica. L'amplificatore differenziale amplifica la differenza di tensione, presentando in uscita una continua. L'utilizzo, nel contesto del circuito riscaldante, di un anello di controreazione, permette di ridurre gli errori, prendendo la continua anche quando questa tenderà a 0.

Questo tipo di meccanismo è molto interessante, ma anche problematico: il fatto di lavorare con segnali così piccoli, implica la necessità di introdurre, in ingresso al convertitore, un amplificatore / attenuatore, al fine di avere una $v(t)$ amplificata e quindi una tensione tangibile in uscita all'amplificatore differenziale.

Introdurre un amplificatore equivale ad introdurre un filtro passa basso, e quindi bisognerà far sì che l'amplificatore contenga tutte le armoniche del segnale che intendiamo studiare. Ciò introduce i problemi di dinamica dell'amplificatore, nel nostro circuito, aumentando la sua inerzia. Introducendo ad esempio un segnale di tipo impulsivo, quindi con V_{eff} basso ma V_P, ossia tensione di picco, molto elevata, l'amplificatore potrebbe saturare in quanto non in grado di rappresentare un segnale a frequenza molto elevata, e dunque falsare il valore efficace presentato dallo strumento. Questa situazione per il nostro voltmetro, che non sarà in grado di presentare dunque una certa gamma di segnali, a causa di questo problema dell'amplificatore; l'unica soluzione, è quella di partire con un'attenuazione più elevata possibile, riducendola man mano, in modo da ridurre l'impulsività del segnale, eliminare il problema della rilevazione del picco troppo elevato, e cercando così di raggiungere un risultato buono, seppur con risoluzione probabilmente non molto elevata.

Capitolo 4

Misure di frequenza

Potremmo incominciar a studiare le misure di frequenza, chiedendoci per quale motivo ci si concentra a misurare una grandezza di questo tipo; la risposta è semplice: le misure di frequenza sono molto semplici da realizzare, e con sistemi poco elaborati si ottengono incertezze relative dell'ordine di 10^{-6}; con sistemi elaborati, si riesce a migliorare di molto, rispetto a questo ordine di grandezza, l'accuratezza e l'incertezza relativa della misura.

Molte grandezze, misurabili in modo diretto, possono essere convertite in frequenze mediante trasduttori, e quindi essere misurate mediante una misura indiretta, basata sulla misura di frequenza, ottenendo un'ottima accuratezza.

Esistono diverse definizioni, legate al concetto di frequenza, ossia diversi tipi di frequenza potenzialmente misurabili; presentiamo un esempio pratico, in grado di permetterci di comprendere meglio queste definizioni. Supponiamo di disporre di un fenomeno ciclico, nella fatispecie di una sinusoide, la cui fase $\psi(t)$ è variabile nel tempo. Se la sinusoide ha velocità costante nel tempo, allora $\psi(t)$ sarà una retta, poichè al variare del tempo si avrà una variazione lineare della fase. Può capitare che $\psi(t)$ sia una curva generica, che segue la retta solo come curva inviluppo; si definisce, a partire dalla funzione $\psi(t)$, il concetto di frequenza istantanea f_i come:

$$f_i = \frac{d\psi(t)}{dt}$$

Partiamo da questa definizione; supponiamo di considerare, anzichè un certo punto t_0 nel quale calcolare la pendenza della curva, ossia la frequenza istantanea, un certo intervallo di tempo ΔT, ossia una differenza di angolo, una differenza tra due punti della fase. Supponiamo di considerare, di tutte le differenze di angolo, una ampia T_1; si definisce dunque, a partire dalla definizione di prima (estesa però in un concetto di media in un dato intervallo di tempo, e non più considerata in un contesto istantaneo):

$$f_m = \left. \frac{\Delta\psi(t)}{\Delta T} \right|_{\Delta T = T_1}$$

Se il periodo è molto grande, potremo pensare di aver ottenuto la pulsazione media del segnale di fase $\psi(t)$ su tutto il dominio del segnale.

A seconda dei metodi di misura, e di cosa vogliamo misurare, potremo studiare frequenze istantanee piuttosto che medie su periodi piuttosto che medie assolute; dal momento che per ora ci concentreremo prevalentemente su misure di frequenza basate su metodi di conteggio, avremo bisogno di utilizzare misure di pulsazioni (di frequenze) medie su periodi.

4.1 Misure di frequenza mediante tecniche di conteggio

Per realizzare una misura di frequenza mediante tecniche di conteggio, ossia una misura in grado di contare il numero di cicli (o periodi) che si ripetono in un certo intervallo di tempo campione, servono sostanzialmente tre ingredienti:

- Un dispositivo in grado di evidenziare elementi ciclici, ossia elementi in grado di evidenziare gli eventi da contare;
- Un dispositivo in grado di contare gli eventi rilevati;
- Un dispositivo in grado di generare una base tempi, ossia una sorta di dominio del tempo discreto, sul quale ci baseremo per misurare frequenze.

4.1.1 Misura diretta di frequenza

Abbiamo un certo numero di impulsi di periodi t_x, che faccio rilevare ed inviare ad un cirucito contatore, il quale definisce una certa durata T_c durante la quale raccoglie eventi, e li conteggia. Supponendo che esso raccolga un numero pari a n di conteggi, avremo che:

$$T_c \simeq n t_x$$

Definita dunque la frequenza incognita degli impulsi che intendiamo contare, f_x, come reciproco di t_x, si avrà che:

$$f_x = \frac{1}{t_x} \simeq \frac{n}{T_c}$$

Il fatto di non avere l'esatta uguaglianza deriva dal fatto che n potrebbe non essere esattamente il numero di eventi rilevati, poichè il contatore potrebbe averne contato uno in meno.

Come si realizza nella pratica tutto ciò? Serve un generatore di tempi campione T_c: esso sarà fondamentalmente composto da un oscillatore al quarzo, campione con accuratezze piuttosto elevate, per prezzi piuttosto bassi. Il quarzo è un elemento piezoelettrico: se eccitato da segnali, si comprime, e quindi genera impulsi elettrici con un'elevata frequenza, lavorando in frequenza di risonanza meccanica. Mediante un circuito di divisione (quale per esempio un contatore) si riesce facilmente a dividere la frequenza del quarzo, f_c, in un suo sottomultiplo (aumentando dunque il periodo T_c nel periodo finale T_c che utilizzeremo per la misura): infatti, se possiamo dividere la frequenza per un fattore k, possiamo ottenere:

$$f_c \Longrightarrow \frac{f_c}{k} \Longrightarrow T_c = kT_c$$

A questo punto, abbiamo un periodo T_c sufficientemente elevato sul quale lavorare; possiamo inviarlo ad un generatore di porta, realizzando il periodo campione sul quale opererà il contatore in grado di contare eventi esterni.

Il problema già citato a questo punto è avere, come rapporto tra il periodo tra due impulsi t_x, e T_c tempo campione da noi introdotto, un numero n intero:

$$n = \frac{T_c}{t_x}$$

Infatti, t_x e T_c sono generati uno da un fenomeno esterno che stiamo analizzando, l'altro dal quarzo, e dunque non presentano alcuna correlazione tra loro: potremmo sbagliare di un impulso in positivo o in negativo. La risoluzione avrà come ultima cifra percepibile 1, e, poichè possiamo sbagliare di ± 1, se includiamo sia il primo che l'ultimo impulso (+), o se non includiamo nè il primo nè l'ultimo impulso (-), la risoluzione massima sarà la minima variazione percepibile, ossia di n, un singolo valore intero, 1; vediamo meglio che:

$$f_x = \frac{n}{T_c}; \quad \Delta f_x = \frac{1}{T_c}$$

La risoluzione relativa, ossia la minor variazione leggibile sullo strumento, sarà:

$$\frac{\Delta f_x}{f_x} = \frac{1}{n}$$

Si noti che non abbiamo ancora parlato di errore, ma solo di risoluzione, ossia di minima variazione percepibile dallo strumento. Per migliorare la risoluzione, riuscendo a rendere percepibile quindi una cifra sempre meno significativa, la tattica che potremo effettuare sarà quella di aumentare n, e per far ciò considerare periodi T_c molto lunghi.

La tecnica sinora esposta, come si potrebbe intuire, è idonea alla misura di frequenze elevate: dato un certo T_c, nel quale si conta un numero di eventi n piuttosto elevato (ottenendo dunque una risoluzione bassa), potremo distinguere chiaramente la frequenza, poichè la durata dei t_x si può considerare trascurabile rispetto a T_c. Purtroppo non sempre ciò è verificabile: per questo motivo, esiste un'alternativa al metodo di misura finora esposta.

4.1.2 Misura diretta di periodo

Se la frequenza f_x del fenomeno che vogliamo studiare è bassa, avremo un problema non trascurabile: $f_x \sim T_c$, dunque la risoluzione diminuisce notevolmente, poichè n diventa sempre più piccolo: se il periodo del fenomeno è infatti dello stesso ordine di grandezza del periodo di misura, non potremo contare molti campioni, e a partire da questi utilizzare le formule da noi introdotte, poichè otterremmo risultati molto poco significativi. Quella che si effettua è una misura del tutto duale alla precedente, ossia una misura diretta di periodo. Essa è duale nel senso che ora il nostro generatore di base tempi al quarzo genererà sì tempi, ma non più un periodo campione T_c, bensì tempi di durata infinitesima, T_c, che ci serviranno per quantificare la durata del periodo incognito, T_x.

T_x sarà il segnale che deciderà la durata del conteggio dei segnali T_c, che supponiamo essere m; T_x contiene diversi t_x base, ossia diversi eventi che insieme formano il periodo T_x: il loro numero si può considerare noto, M: esso sarà una costante introdotta dal sistema, poichè, per ottenere un periodo T_x sufficientemente lungo, attenderemo il passaggio di M elementi t_x. Fatte le ipotesi preliminari e la presentazione delle convenzioni usate, possiamo dire che:

$$T_x \simeq mT_c$$

$$T_x = Mt_x$$

$$\Longrightarrow t_x = \frac{m}{M}T_c$$

A seconda della frequenza del segnale o fenomeno ciclico che vogliamo quantificare, conoscendo anche solo la differenza di ordine di grandezza tra le due, sarà possibile scegliere senza dubbi quale dei due metodi utilizzare: questo ci permetterà di ottenere una risoluzione migliore, ai fini di non avere un collo di bottiglia al momento del calcolo dell'incertezza.

Esiste, proprio per quanto riguarda la risoluzione, un diagramma in grado di suggerirci come operare, al momento di fare una scelta di questo tipo:

A seconda della risoluzione desiderata, e dell'ordine di grandezza del segnale da misurare, si possono quindi trarre suggerimenti sul metodo di misura da applicare.

Si vuole ricordare che fino a qua si sono introdotti due differenti metodi, per poter migliorare la sola risoluzione del sistema di misura; ai fini di poter ottenere maggior accuratezza delle misure, ciò non è purtroppo sufficiente.

4.1.3 Incertezze di misura

Possiamo dire che certamente la risoluzione fa parte di quello che potremmo definire il budget delle incertezze; al fine di determinare l'accuratezza dello strumento, tuttavia, siamo molto lontani dal risultato finale che vorremmo determinare. Abbiamo analizzato due casistiche, due metodi di misura duali: alcuni contributi di incertezza si faranno sentire sia nel caso di misura di frequenza che in quello di misura di periodo. Avremo bisogno di introdurre per i singoli casi però alcune leggere varianti, come vedremo tra breve.

Ricordiamo le nostre relazioni, per quanto riguarda la misura di frequenza:

$$f_x = \frac{1}{t_x} \simeq \frac{n}{T_c}; \quad T_c = kT_c$$

La variazione relativa di frequenza, dunque, avrà un'espressione del tipo:

$$\varepsilon_{f_x} = \frac{\delta f_x}{f_x} = -\frac{\delta T_c}{T_c} + \frac{\delta n}{n}$$

Dal momento che il moltiplicatore (o divisore) k è un parametro determinato, introdotto da noi e privo di errore:

$$\frac{\delta T_c}{T_c} = \frac{\delta t_c}{t_c}$$

Abbiamo dunque che:

$$\varepsilon_{f_x} = -\frac{\delta t_c}{t_c} + \frac{\delta n}{n}$$

Sappiamo che la frequenza campione, f_c, si definisce come reciproco del tempo campione, t_c; l'incertezza relativa, dunque, potrà ricondursi all'incertezza del tempo campione:

$$f_c = \frac{1}{t_c} \Longrightarrow \frac{\delta f_c}{f_c} = -\frac{\delta t_c}{t_c}$$

Utilizzando modelli di vario tipo, sapendo che i contributi all'incertezza sono questi, possiamo quantificare quantomeno questa prima parte dell'incertezza. Analizziamo più nei dettagli questi contributi

Incertezza di quantizzazione

Il tempo è un elemento variabile con continuità, nel mondo reale; il fatto che introduciamo una base tempi, ossia il fatto che discretizziamo una quantità continua, introduce un'indeterminazione nel nostro sistema. Quelli che misuriamo, infatti, non sono esattamente gli eventi, ma una parte di essi in grado di identificarli, in qualche modo:

Dobbiamo essere in grado di calcolare il numero di fenomeni avvenuti contando esattamente il numero di trigger (per esempio i fronti in salita di un impulso). Può capitare che, a causa di una cattiva scelta del periodo di funzionamento (tempo campione T_C), o per il fatto che in esso non siamo in grado di comprendere esattamente il numero di eventi, senza contarne uno in meno o uno in più, si abbia per l'appunto un errore di quantizzazione. Per questo motivo, dal momento che possiamo sbagliare di un campione su n, $\delta n = 1$, e quindi l'incertezza relativa introdotta dall'errore di quantizzazione sarà:

$$\frac{\delta n}{n} = \frac{1}{n}$$

Incertezza del campione al quarzo

La base tempi è determinata da un dispositivo al quarzo, che come già detto, sfruttando le proprie proprietà piezoelettriche, è in grado di fornire un campione di frequenza abbastanza accurato. Dal momento che la temperatura può essere un elemento in grado di variare il volume di un oggetto, ed essendo la piezoelettricità basata proprio sulla compressione dell'elemento al quarzo, termostatizzando l'elemento si riesce ad ottenere anche un'accuratezza relativa di questo tipo:

$$\frac{\delta f_c}{f_c} = \frac{\delta t_c}{t_c} \simeq 10^{-8} \div 10^{-9}$$

L'errore di quantizzazione si può dunque ridurre aumentando di molto la lunghezza del periodo T_c, e dunque il numero n di campioni acquisiti dal sistema contatore, ma l'incertezza del campione al quarzo è intrinseca, e oltre alla termostatizzazione non si può effettuare alcuna operazione in grado di migliorare l'accuratezza.

Piccola osservazione: la risoluzione coincide con la sola incertezza relativa di quantizzazione; poichè il campione al quarzo introduce un'ulteriore indeterminazione, non è detto che tutte le cifre dello schermo siano significative (a meno che la risoluzione non sia sufficientemente bassa da non essere influenzata dall'incertezza della base tempi al quarzo).

Incertezza aggiuntiva sulle misure di periodo

Per le misure dirette di frequenza, questo è quanto; se affrontiamo una misura della seconda categoria, ossia una misura di periodo, dovremo introdurre, oltre a tutti i termini sinora analizzati, un ulteriore termine di indeterminazione: il misurando, in misure di questo ambito, è il tempo T_x di durata del conteggio. L'incertezza aggiuntiva è sul tempo di inizio e tempo di fine del periodo di gate, ossia della durata del conteggio: quest'indeterminazione deriva da contributi di rumore, che si vanno ad aggiungere sul segnale, introducendo un'ulteriore incertezza. Se prima, nella misura diretta di frequenza, non avevamo problemi di questo tipo in quanto il tempo T_c era perfettamente determinato, una volta risolti i problemi legati al quarzo, ora non possiamo dire lo stesso.

Il rumore può comportare, in diversi punti della misura, conteggi in più, causati dalla presenza di tempi di errore t_e, che non dovrebbero esserci, ma invece esistono a causa di errate commutazioni dei comparatori contenuti all'interno dei contatori, causate da un rumore di ampiezza V_n. Si può calcolare che:

$$\frac{V_n}{t_e} = \frac{dv_x}{dt}$$

In termini relativa, l'incertezza introdotta sul periodo misurando T_x da contributi di errore è:

$$\varepsilon_{T_x} = \frac{2t_e}{T_x}$$

Al solito, nel caso più completo e di worst case, l'incertezza risultante sulla misura sarà data dalla somma in modulo di tutti i contributi di errore finora introdotti.

4.1.4 Misura di intervalli di tempo

Caso particolare di misura di periodi, serve a quantificare il ritardo, la distanza temporale tra due eventi singoli. Significa, in termini tecnici, misurare l'intervallo di tempo compreso tra un evento di start ed un evento di stop. L'analogia con la misura di periodo sembra evidente, ma non bisogna lasciarsi troppo trarre in inganno: il periodo, l'intervallo di tempo misurato, è unico: non si ha media, e non si hanno eventi ripetuti, ma i soli eventi di start e di stop. L'intervallo di tempo incognito, T_0, dato n il risultato di conteggio, e t_c il tempo campione:

$$T_0 = nt_c$$

Risoluzione di lettura

Lo studio della risoluzione di lettura non ci dirà niente di nuovo: poichè sono valide le equazioni prima presentate, e poichè al minimo possiamo contare un tempo t_c o uno in meno:

$$\Delta T_0 = t_c = \frac{1}{f_c}$$

La risoluzione relativa sarà, come al solito:

$$\frac{\Delta T_0}{T_0} = \frac{t_c}{nt_c} = \frac{1}{n}$$

Incertezza della misura

Per quanto riguarda lo studio dell'incertezza, dovremo affrontare sostanzialmente tre tipi di indeterminazione:

- L'incertezza di quantizzazione, definita esattamente come negli altri casi finora analizzati;
- L'incertezza introdotta dal quarzo, anch'essa del tutto analoga ai casi finora analizzati;
- L'incertezza sul tempo di start e di stop, dovuta al rumore presente sul segnale e sul fenomeno analizzato, con l'incertezza relativa precedentemente mostrata.

L'incertezza relativa risultante dalla combinazione degli elementi appena presentati, sarà:

$$\frac{\delta T_0}{T_0} = \frac{1}{n} + \frac{\delta f_c}{f_c} + 2\frac{t_e}{T_0}$$

4.2 Voltmetri numerici (DVM)

I voltmetri numerici (o DVM: Digital Volt Meter) sono dispositivi in grado di effettuare misure numeriche di tensioni: data una tensione, un trasduttore analogico/digitale la converte in una sequenza di numeri, che verran poi elaborati da un frequenzimetro (come vedremo meglio in seguito). Analizzeremo nel dettaglio voltmetri numerici (o digitali) a semplice e doppia integrazione: in realtà ora in commercio si trovano solo dispositivi a doppia integrazione, o addirittura a integrazione multipla, tuttavia conoscere l'idea di base può risultare sempre utile.

I voltmetri digitali presentano una misura di tensione su di uno schermo digitale; la cosa interessante, come si sarà capito dall'introduzione, è il fatto che la presentazione non è l'unico elemento numerico presente nel voltmetro. Unico vero vincolo, è il fatto che, intrinsicamente, i voltmetri numerici sono in grado di misurare esclusivamente tensioni continue; introducendo circuiti in grado di modificare l'alternata esprimendola in termini di una continua equivalente (quali per esempio i circuiti utilizzati per realizzare voltmetri analogici a valore efficace) è possibile elaborare quest'ultima, e quindi utilizzare il DVM anche in un ambito di grandezze variabili nel tempo. Ovviamente, utilizzando le leggi di Ohm, ed alcuni accorgimenti, è possibile utilizzare questo tipo di strumento da multimetro, misurando dunque anche correnti, o resistenze.

Alla base del voltmetro analogico, come già suggerito, è un convertitore A/D (analogico/digitale), ossia un dispositivo in grado di trasformare un segnale analogico in un numero. Questo numero, codificato in una certa maniera, sarà in grado di esprimere la tensione di uscita, che verrà presentata alla periferica di output. L'attivazione del dispositivo darà luogo ad un processo di conversione, che durerà un certo tempo, detto tempo di conversione; terminato questo tempo, avremo ottenuto il risultato numerico N esprimente il valore della tensione misurata. Poichè scegliamo di utilizzare una misura di frequenza al fine di misurare indirettamente una tensione, il numero che dovremo interpretare come tensione dovrà essere molto accurato, ossia dovremo essere in grado di ottenere un'incertezza ridotta. In ambito di DVM, dunque, i convertitori A/D potranno essere anche lenti, ma dovranno essere accurati e in grado di rimuovere disturbi in maniera efficace.

Come si è capito da questo discorso, si possono avere sostanzialmente due tipi di misurando:

- Tensioni alternate, AC;
- Tensioni continue, DC.

Tensioni alternate

In sostanza il DVM è in grado di misurare esclusivamente tensioni continue; quello che dobbiamo fare è cercar di campionare un valore di tensione alternata, utilizzando un campionatore, S/H (Sample and Hold):

Un impulso di campionamento, di Sample, chiude il circuito per un istante, durante il quale il condensatore, elemento con memoria, si carica; dal momento che termina l'istante di Sample, si passa alla fase di Hold: il circuito è aperto, e dunque il condensatore, carico, mantiene la propria carica, che potrà essere interpretata in tensione utilizzando le equazioni di funzionamento del condensatore.

Tensioni continue

La tensione continua in realtà presenta diversi tipi di disturbi, uno su tutti quello della rete elettrica; mediante alcuni accorgimenti, che vedremo in seguito, sarà possibile eliminarli con furbizia.

Terminato questo cappello introduttivo, entriamo nel vivo dell'argomento, passando alla descrizione dei voltmetri basati sul principio di integrazione.

4.2.1 Voltmetri a integrazione semplice

Abbiamo già citato il nome integrazione, indicando il tipo di voltmetri digitali che tratteremo; cerchiamo di meglio chiarire il concetto dietro a questo nome, al fine di capire come progettare un voltmetro a singola integrazione, e poi evolverci.

Il concetto è semplice: un circuito come quello descritto in precedenza, parlando di S/H: un convertitore tensione/frequenza, basato sulla carica di un condensatore con una corrente prodotta dalla tensione (costante) incognita; una volta caricato del tutto il condensatore, raggiunta nella fatispecie una certa soglia, si inizia a svuotare con uno svuotamento costante (tenendo conto che nel mentre continua anche a riempirsi, con il principio di prima; lo svuotamento prevale comunque sul riempimento, altrimenti non avrebbe senso). Una volta tornati sotto soglia, e raggiunto di nuovo il livello regolare, si stoppa il processo di svuotamento; quando ricaricato si ri-riempie, poi si ri-svuota, e così via.

All'inizio il circuito integratore integra solo la V_x, che però passa dal comparatore di soglia, il quale rileva il valore integrato; questo viene retroazionato e sommato a V_x, in modo da continuar a integrare una quantità sempre maggiore; il comparatore di soglia ad un certo punto sentirà il fatto che la tensione ha caricato troppo il circuito, e quindi inizierà a far sì da svuotare il condensatore.

Possiamo provare a interpretare come segue il tutto: l'integratore, cuore del circuito, riempie una sorta di vasca, rappresentata dal condensatore, ed il comparatore di soglia, guardando quanto la vasca è piena, ordina di portar via dell'acqua e gettarla fuori dalla vasca (nel nostro circuito, svuotare il condensatore togliendovi carica).

Si noti che l'impulso scaricante il condensatore ha ampiezza e durata a noi noti, poichè queste sono caratteristiche intrinsiche dei dispositivi utilizzati nel circuito; definiamo ora le grandezze, al fine di poter quantificare in modo più formale tutto ciò che abbiamo detto.

Il processo di integrazione dura per un periodo T, durante il quale si continua in pratica ad integrare (come nella normale operazione di integrazione) la tensione V_x (utilizzando il circuito prima descritto). Dopo il periodo T, periodo che il condensatore impiega per caricarsi, e quindi istante in cui il comparatore di soglia si accorge dello stato del condensatore, parte un impulso di durata T_0 ed ampiezza E_0 che sottrae tutta la carica introdotta nel condensatore fino a riportarlo allo stato iniziale.

Poichè il circuito è basato su di un amplificatore operazionale, sappiamo che ha due morsetti di ingresso, e che ha un guadagno $A \rightarrow +\infty$; la tensione tra il morsetto di $+$ e quello di $-$ è vicina a 0 V; l'impedenza in ingresso, ossia l'impedenza vista dai morsetti di ingresso, tende a infinito. Si può dunque creare una maglia, che è interrotta: tra i due morsetti di ingresso infatti c'è un taglio, ma al contempo c'è una differenza di tensione nulla; possiamo dunque dire che:

$$I_R = \frac{V_x}{R}$$

La corrente I_R va dunque tutta nel condensatore, ed è costante. La tensione ai capi di C, per definizione di tensione ai capi del condensatore, sarà pari a:

$$V_C(t) = \frac{Q(t)}{C} = \frac{1}{C}\int_T I_R(t)dt = \frac{1}{C}\int_T \frac{V_x}{R}dt = frac1RC \int_T V_x dt$$

Il comparatore di soglia è basato, come l'integratore, sull'uso di un operazionale; in un operazionale, si hanno due zone di saturazione, separate idealmente da una discontinuità tipo salto, la cui pendenza infinita è giustificata dal guadagno $A \to \infty$. Quando dunque si ha l'evento di commutazione, il comparatore salta, e questo salto permette di inviare un impulso indicante il fatto che la tnesione abbia raggiunto una certa soglia.

Possiamo dunque scrivere che:

$$\frac{1}{R_1 C}\int_T V_x dt = \frac{1}{R_2 C}\int_{T_0} E_0 dt$$

Poichè i DVM come già detto lavorano solo con tensioni continue, V_x sarà una costante rispetto al tempo t; E_0 invece è l'impulso da noi stabilito in grado di svuotare il condensatore, e di fatto si può rappresentare come un'onda quadra la cui parte alta ha durata pari a T_0. Poichè anche E_0 è costante, dunque, possiamo integrare, ottenendo:

$$\frac{V_x}{R_1}T = \frac{E_0}{R_2}T_0$$

Modificando leggermente questa formula, possiamo ricavare un'espressione mediante la quale è possibile misurare indirettamente la tensione V_x:

$$V_x = E_0 \frac{T_0}{T}\frac{R_1}{R_2} = E_0 T_0 \frac{R_1}{R_2} f$$

Dove la frequenza f identifica la durata degli impulsi T_0:

$$f = \frac{1}{T}$$

Si suol semplificare, introducendo un termine K come costante strumentale:

$$E_0 T_0 \frac{R_1}{R_2} \triangleq K$$

Incertezze di misura del voltmetro ad integrazione semplice

Poichè disponiamo della funzione:

$$V_x = g(f)$$

Possiamo, a partire da essa, ricavare i contributi di errore del nostro DVM. Possiamo catalogarli, dividendo dunque le fonti di incertezza in diversi contributi:

- Incertezza sull'ampiezza dell'impulso di scarica, E_0:

$$\frac{\delta E_0}{E_0}$$

- Incertezza sulla durata dell'impulso di scarica, T_0:

$$\frac{\delta T_0}{T_0}$$

- Incertezza sulle resistenze con le quali si costruisce il circuito: si noti che queste resistenze, molto spesso, sono matched (ossia unite): si può considerare di esse direttamente l'incertezza del rapportro delle resistenze, anzichè due contributi di incertezza separati, poichè possono essere fornite con valori di incertezza tra loro correlati;

- Incertezza sulla frequenza dei campioni; essa dipende sia dall'accuratezza con la quale sono definiti i campioni, sia dal numero di conteggi (errore di quantizzazione intrinseco del contatore):

$$\frac{\delta f}{f} = \frac{\delta f_c}{f_c} + \frac{1}{n}$$

Fino a qua abbiamo considerato cause di incertezza non consideranti eventuali non idealità del circuito realizzato; alcune cause di non idealità del circuito, introducenti un ulteriore contributo di incertezza, sono:

- Eventuali offset sulla tensione prodotta dall'integratore: l'integratore è basato come sappiamo su di un operazionale; se l'operazionale è ideale, la tensione tra i morsetti è 0 V; poichè il mondo non è però ideale, la tensione tra il $+$ ed il $-$ è non nulla, e quindi questo offset può creare incertezza sui nostri risultati;

- Offset sui comparatori di soglia: eventuali offset dei comparatori di soglia, introducenti incertezza dunque sull'offset, non possono cambiare la frequenza, ossia non introducono, sulla frequenza, un'ulteriore incertezza;

- Cattiva definizione della tensione di soglia del comparatore: come il secondo caso, non introduce un errore sulla frequenza.

Reiezione del rumore

Un effetto finora non considerato che potrebbe effettivamente provocare un notevole disturbo, è il rumore presente sulla soglia: esso potrebbe effettuare una modulazione sul periodo degli impulsi, e quindi variare la frequenza. Cerchiamo di capire se c'è del rumore, e che contributi esattamente ci fornisce: un rumore, su di una misura media di periodo, fornisce un contributo pari alla sua area mediata nel periodo T_m, e dunque fornisce un contributo pari alla sua media, alla sua continua. Una volta valutato il tipo di rumore, e l'area dell'impulso, possiamo cercar di effettuare una reiezione del rumore dalla nostra misura, in modo da eliminare questo tipo di incertezza.

Se abbiamo a che fare con un rumore di tipo impulsivo, ossia che colpisce, nel dominio del tempo, solo singoli attimi di tutto il segnale, allora, data A la sua ampiezza, e T_m il tempo sul quale si effettua la misura media di periodo del segnale, la componente continua equivalente al rumore sarà semplicemente:

$$\frac{A}{T_m}$$

Non esistono però solo rumori di tipo impulsivo: esistono rumori in un certo senso determinati, in grado di mantenere una certa ciclicità: esempio di questi sono i rumori sinusoidali. La cosa interessante è che però le sinusoidi sono sostanzialmente segnali a media nulla, su di un loro periodo; se siamo capaci di trovare una condizione in grado di sfruttare questa cosa, possiamo eliminare il rumore introdotto da segnali di questo tipo. Dovremo lavorare, al fine di ridurre il rumore, sul tempo misurato T_m, relazionandolo con la frequenza del segnale di rumore. Possiamo intuire facilmente che, avendo a che fare con sinusoidi, il valor medio più elevato si otterrà se il T_m si fermerà a metà del periodo di un segnale di rumore sinusoidale: in questo modo, la media del rumore sarà massima, e quindi introdurrà un grosso discostamento da ciò che vorremmo misurare. Data sinusoide di ampieza A_{sc}, e periodo T_d:

$$T_m = (2k+1)\frac{T_d}{2}$$

Sarà senza dubbio il caso di worst case.

Quello che dovremo fare, è utilizzare come tempo di misura T_m un multiplo del periodo del rumore T_d; tra poco esporremo un esempio pratico, dopo aver introdotto un formalismo in grado di quantificare la reiezione al disturbo: il nostro scopo è valutare l'area del rumore, calcolarne la continua equivalente, e cercare di eliminarla. Possiamo utilizzare quello che in analisi si chiama teorema della media integrale, e mediante esso definire la caratteristica di reiezione al disturbo come:

$$R_N = V_0 \left[\frac{1}{T_m} \int_{T_m} V_d \sin(2\pi f_d t) dt \right]$$

Dove f_d si definisce come il reciproco del periodo di rumore, T_d.

Supponiamo di aver a che fare ad esempio con il più classico dei disturbi: quello di rete (poichè la rete elettrica è alimentata mediante un'alternata a 50 Hz); volendo eliminare ad esempio il disturbo di rete, dovremo lavorare con un multiplo dell'inverso di 50 Hz, e quindi almeno con:

$$\frac{1}{50\ Hz} = 20\ ms$$

Il tempo T_m dovrà essere pari a 20 ms o ad un suo multiplo.

4.2.2 Voltmetri numerici a doppia integrazione

I voltmetri numerici a integrazione doppia rappresentano un'evoluzione rispetto ai precedenti a singola integrazione, anche se c'è qualche differenza sulle idee di fondo sfruttate dall'uno e dall'altro: se prima convertivamo una tensione in una frequenza, ora, sempre mediante il principio di integrazione (ed un circuito un po' più complicato), convertiamo una tensione in un intervallo di tempo, misurato mediante il conteggio degli impulsi di un oscillatore al quarzo.

Il principio di funzionamento è il seguente: consideriamo il processo di misura diviso in due sostanziali fasi, dove effettueremo due operazioni del tutto duali tra loro:

- In un primo tempo integriamo una tensione a noi incognita, V_x, per un certo tempo campione, T_c; questo lo si farà considerando anche la costante di tempo introdotta dal circuito, $\tau = RC$, ottenendo dunque una quantità di carica proporzionale alla V_x:

$$Q(T_c) = \frac{1}{RC} \int_{T_c} V_x dt$$

- Una volta trascorso il periodo T_c, effettuiamo un'operazione duale: a partire dal valore di carica $Q(T_c)$, che sarà stato accumulato in un condensatore, introduciamo nel circuito una V_r, ossia una tensione di riferimento a noi nota con polarità inversa alla V_x; dopo un certo periodo T_x, che dovremo misurare, l'effetto dell'integrazione della V_r avrà portato a 0 il numero di cariche contenute all'interno del condensatore. La durata del tempo di scarica, T_x, sarà proporzionale alla carica che la tensione incognita V_x ha introdotto nel circuito durante il tempo campione T_c, e dunque potremo misurare indirettamente V_x; il processo di integrazione di V_r vale:

$$\frac{1}{RC}\int_{T_x} V_r dt$$

Da queste ipotesi preliminari, possiamo capire semplicemente che l'integrazione della tensione incognita equivale all'integrazione della tensione nota, dal momento che il risultato finale è quello di partire da una carica nulla e tornare ad una carica nulla. Possiamo dire, poichè sia V_x che V_r sono tensioni costanti, che:

$$\frac{1}{RC}\int_{T_c} V_x dt = \frac{V_x}{RC}T_c$$

$$\frac{1}{RC}\int_{T_x} V_r dt = \frac{V_r}{RC}T_x$$

Poichè le due sono uguali tra loro:

$$\frac{V_x}{RC}T_c = \frac{V_r}{RC}T_x \Longrightarrow V_x = V_r\frac{T_x}{T_c}$$

Lo schema a blocchi non sarà così diverso dal precedente, se non un po' più elaborato.

Il circuito è realizzato in modo da non far intervenire la costante di tempo τ; in questo modo, i calcoli si semplificano notevolmente. La tecnica di misura del periodo è quella di misura diretta, analizzata precedentemente parlando di misure di frequenza; dati N_x il numero di impulsi contati durante il periodo incognito dal nostro oscillatore al quarzo, e N_c il numero di impulsi contati durante il periodo noto, dal momento che:

$$T_c \simeq N_c t_c; \quad T_x \simeq N_x t_c$$

Dove t_c è il periodo base dell'oscillatore al quarzo, allora possiamo dire che:

$$V_x \simeq V_r\frac{N_x}{N_c}$$

Ciò è molto interessante poichè, ai fini della sola determinazione del valore misurato, non ci è necessario conoscere con esattezza la frequenza f_c dell'oscillatore al quarzo, dal momento che siamo riusciti a ricondurre la misura di tensione V_x semplicemente al conteggio di eventi, senza disporre di altri dati (se non la tensione di riferimento V_r).

Si noti che il fatto che abbiamo potuto eliminare la dipendenza dalla frequenza dell'oscillatore al quarzo f_c nella nostra relazione ci permette di eliminare un termine di incertezza, rendendo già questa tecnica preferibile, in fatto di accuratezza, rispetto a quella basata sull'integrazione semplice; altro contributo di incertezza che scompare, è quello relativo alle costanti di tempo τ, e quindi al resistore R e al condensatore C.

Risoluzione del DVM a doppia integrazione

Riprendendo semplicemente la formula appena introdotta:

$$V_x \simeq V_r \frac{N_x}{N_c}$$

Possiamo quantificare la minima variazione presentabile dallo strumento, ΔV_x, come:

$$\Delta V_x \simeq \frac{|V_r|}{N_c} \simeq \frac{|V_x|}{N_x}$$

Passiamo ora alla stima delle incertezze.

Stima delle incertezze del DVM a doppia integrazione

Dalla solita relazione di partenza,

$$V_x \simeq V_r \frac{N_x}{N_c}$$

Si può ricavare che l'incertezza relativa sulla misura di V_x, stimata mediante un modello di worst case, derivi da:

$$\varepsilon_{V_x} = \left| \frac{\delta V_x}{V_x} \right| = \left| \frac{\delta V_r}{V_r} \right| + \frac{1}{N_c} + \frac{1}{N_x}$$

Ossia consideriamo, oltre ai contributi dell'errore dettato dalla cattiva determinazione del campione V_r, anche gli errori di quantizzazione introdotti dall'operazione di contare gli impulsi. Possiamo in effetti fare di meglio: poichè la prima parte del tempo di integrazione (T_c) è nota, possiamo scegliere di sincronizzare l'avvio dei conteggi con lo start dell'integrazione, in modo da eliminare di fatto l'indeterminazione su N_c. Di fatto, quindi, potremo eliminare il termine relativo a N_c, ottenendo semplicemente:

$$\varepsilon_{V_x} = \left| \frac{\delta V_x}{V_x} \right| = \left| \frac{\delta V_r}{V_r} \right| + \frac{1}{N_x}$$

Da un'analisi circuitale, come nel caso di voltmetro a singola integrazione, è possibile determinare un insieme di altri parametri in grado di aumentare l'incertezza dello strumento. Questi sostanzialmente sono del tutto identici a quelli prima introdotti nel caso di integrazione semplice, più uno extra: poichè questo tipo di circuito è più elaborato rispetto ai precedenti, potrebbe farsi sentire in misura maggiore la sua inerzia, la sua latenza, e quindi la non idealità dei componenti elettrici potrebbe introdurre ritardi tra operazioni che dovrebbero in realtà essere sincronizzate.

Reiezione del rumore

Come anche il voltmetro a singola integraziome anche in questo ambito si può aver a che fare con rumori, in grado di aumentare l'indeterminazione della misura. In realtà però rispetto a questo caso non c'è molto da aggiungere: ai fini di introdurre la reiezione, è sufficiente far sì che il tempo di misurazione, $T_m = T_c + T_x$, sia un multiplo del periodo del disturbo, T_d.

4.2.3 Specifiche dichiarate di un DVM

Nel datasheet o manuali di un DVM, di solito viene dichiarato un certo numero di informazioni, legate a sue caratteristiche. Conoscere il significato di ciò, permette di avere facilitazioni nell'interpretare il risultato ce viene fornito dal DVM in seguito al processo di misura.

- Numero di cifre: si indica il numero di cifre, con una notazione di questo tipo: $3\frac{1}{2}$; $4\frac{1}{2}$, e simili: in questo ambito, vuol dire che abbiamo 3 o 4 cifre a variazione piena, ossia in grado di assumere tutti i numeri da 0 a 9, e la cifra di peso più elevato può assumere solo alcuni valori ($0 \div 1$, piuttosto che $0 \div 2$, o simili, ma non $0 \div 9$, se no sarebbe a variazione piena).

- Fondo scala: il valore di fondo scala (nel caso di un DVM il V_{FS}) è il valore a cui si fa riferimento per parte delle valutazioni dell'incertezza di misura; si tratta del massimo valore esprimibile, ad una certa portata dello strumento;

- Overrange: lettura massima consentita, in %, oltre al valore di fondo scala, rimanendo comunque nella portata di questo. Se abbiamo ad esempio un overrange del 100%, con un fondo scala V_{FS} pari a 1 V, allora potremo leggere fino a 1,999 V (se abbiamo un numero di cifre $3\frac{1}{2}$); se l'overrange alle stesse condizioni fosse del 50%, potremmo leggere fino a 1,499 V.

Accuratezza di un DVM

Esistono sostanzialmente due contributi sull'incertezza di misura di un DVM:

- Contributo a valore assoluto costante: per qualsiasi punto del campo di misura, in una certa portata, si utilizza il valore di fondo scala ai fini del calcolo di questo contributo; ciò racchiude in sè errori di offset e di non linearità dello strumento;
- Contributo a valore relativo costante: a seconda della misura effettuata, in una certa portata, si riutilizza il valore stesso della misura ai fini del calcolo di questo contributo; l'errore che si quantifica in questa maniera deriva da cattive definizioni della funzione di traferimento del DVM, e quindi su errori di guadagno.

Si ricava, mediante il seguente calcolo, l'incertezza assoluta derivante da questi due contributi:

$$|\delta V_x| = |\varepsilon_1 \% V_{FS} + \varepsilon_2 \% V_x|$$

Dove il termine relativo ε_1 è relativo al contributo a valore assoluto costante, e ε_2 al contributo a valore relativo costante.

Capitolo 5

Oscilloscopi Numerici

Con l'avvento dell'era digitale, gli strumenti un tempo analogici tendono a trovare un loro corrispettivo numerico, o digitale: questo perchè gli strumenti di misura (e non solo di misura) tendono sempre più ad avvicinarsi a dei calcolatori. Introdotto un frontend tra analogico e digitale, ossia un'interfaccia in grado di tradurre il mondo analogico in un insieme di numeri, si tende ad analizzare un insieme di numeri semplicemente mediante computazioni effettuate da calcolatori di vario tipo.

L'equivalente di quello che abbiamo finora conosciuto come oscilloscopio analogico, nel mondo digitale, è il DSO, o Digital Storagen Oscilloscope. Esso nasce sia per il trend dell'elettronica ad andare verso il digitale, sia per superare alcuni limiti degli oscilloscopi analogici: per fare un esempio che dopo sarà meglio chiarito, un oscilloscopio analogico non è in grado di analizzare un fenomeno di tipo transitorio, o comunque un generico evento non periodico; l'oscilloscopio digitale supera questo limite, permettendo all'utente di analizzare e lavorare anche su segnali molto più particolari di quelli che potevamo analizzare con l'oscilloscopio analogico.

Sotto il punto di vista del pannello di controllo, sul quale deve lavorare l'operatore della misura, sembrerebbe di aver a che fare con qualcosa di molto vicino al vecchio dispositivo; in realtà uno strumento di questo tipo è molto più complicato da utilizzare (o quantomeno molto diverso) sia dal punto di vista della realizzazione, ossia della componentistica al suo interno, sia sui modi di utilizzo: bisogna aver ben chiari alcuni concetti, affinchè l'oscilloscopio mostri ciò che noi ci aspettiamo di vedere sullo schermo (e non solo, come vedremo). Si possono comunque effettuare le operazioni che facevamo prima con l'oscilloscopio analogico, poichè sostanzialmente la presentazione del risultato finale non è troppo distante da quella del predecessore di questo strumento.

Il principio sul quale si basa l'oscilloscopio numerico è il campionamento: i valori analogici del segnale vengono campionati in un certo numero di istanti, utilizzando diverse modalità di campionamento (che tratteremo in seguito); in questo modo, si ottiene una quantizzazione dei campioni, che vengono quindi interpretati numericamente, ed immagazzinati in una memoria (come suggerisce il nome DSO, uno Storage). Da questa memoria sarà possibile andare a rileggere i dati, elaborarli, e trattarli anche con strumenti diversi dall'oscilloscopio (esportando ad esempio la memoria mediante floppy disk o porte di vario tipo).

Il campionatore continua a lavorare; dal momento che l'utente introduce un segnale di trigger (che meglio discuteremo in seguito) la memoria inizia a caricarsi, di tutti i dati registrati dall'oscilloscopio; mediante operazioni interne a partire dai campioni l'oscilloscopio potrà ricostruire il segnale e rappresentarlo sullo schermo (a raggi catodici o a LCD); noi non vediamo quindi sullo schermo direttamente l'operazione di campionamento, ma solo un'operazione di lettura dalla memoria, ed elaborazione dei dati in modo da presentare il fenomeno, anche se di fatto esso è già passato (poichè di esso leggiamo solo ciò che abbiamo impresso in memoria). La memoria viene riletta ciclicamente, e così l'immagine disegnata continuamente sullo schermo; il fatto di leggere da memoria, e non di disegnare direttamente un fenomeno durante il suo tempo di vita, ci permette di poter analizzare i transitori, e quindi fenomeni non ciclici, poichè di essi conserviamo semplicemente alcuni punti, memorizzandoli.

Tutte queste operazioni di memorizzazione e trasporto vengono gestite da un microprocessore contenuto nel DSO, il quale gestisce al contempo le operazioni di campionamento, di trigger, di elaborazione dati, di presentazione, e di trasporto dei dati acquisiti durante tutte queste operazioni, arbitrando un BUS di dati.

Altra nota riguarda il trigger: esso ha un significato fondamentalmente diverso rispetto a quello dell'oscilloscopio analogico, poichè esso determina non l'inizio della visualizzazione, bensì l'inizio della memorizzazione dei dati; accedendo alle memorie, si potrà scegliere quali dati presentare, effettuando operazioni di pre-trigger e post-trigger (dal momento che le memorie possono contenere dati anche contenuti prima e dopo gli istanti di trigger, se pensiamo al trigger come l'istante di inizio rappresentazione, come nell'ambito dell'oscilloscopio analogico).

Abbiamo finora parlato di soli pregi di questo oggetto, che in realtà presenta alcune limitazioni: in passato, quando l'elettronica digitale non era ancora sviluppatissima, il fatto che si utilizzi come principio di funzionamento il teorema del campionamento, richiedeva soddisfatta la condizione di Nyquist, e quindi la frequenza di campionamento doveva essere quantomeno doppia rispetto alla frequenza con la quale si intendeva rappresentare, in memoria, il segnale. Ciò ci fa capire che fino a qualche tempo fa, non si potevano rappresentare segnali con frequenze superiori a qualche centinaio di MHz. Ora si riescono ad ottenere buone prestazioni anche per frequenze piuttosto elevate, anche se strumenti in grado di farlo sono piuttosto costosi.

5.1 Analisi dei componenti del DSO

Entriamo più nel merito dei componenti del DSO, al fine di comprenderne meglio il funzionamento ed i principi sui quali si basa. Separatamente discuteremo le diverse modalità e proprietà di alcuni componenti in particolare, quali campionatore e altri.

Unità di ingresso

L'unità di ingresso del DSO inizia con un attenuatore tarato, ed un preamplificatore, in grado di regolare in maniera idonea alle caratteristiche interne dell'oscilloscopio il segnale; opzionale, anche se sarebbe buona norma inserirlo, è un filtro anti-aliasing: esso serve a limitare in banda il segnale, in modo da evitare di ottenere sovrapposizioni di code di altri campioni, al momento dell'operazione di discretizzazione. Normalmente, un filtro di questo tipo, è semplicemente un passa-basso.

Campionatore

Il campionatore, le cui modalità di funzionamento saranno presentate in seguito, è composto sostanzialmente dal solito blocco S/H (Sample/Hold), in grado di campionare alcuni punti del segnale, e convertirli, mediante un trasduttore A/D, in numeri. Il campionatore è regolato da un clock, realizzato mediante (ad esempio) un quarzo, il quale detta i tempi sia del campionatore S/H, sia del convertitore A/D, sia della CPU, ossia del microprocessore in grado di regolare tutte le operazioni: i campioni numerici, in uscita da A/D, infatti, andranno a memorizzarsi nel blocco che ora tratteremo:

Memorie

Le memorie utilizzate per lo storage dei dati dell'oscilloscopio numerico sono sostanzialmente delle RAM, ossia memorie ad accesso casuale; esse sono organizzate secondo una struttura FIFO (First In First Out): il primo campione introdotto in memoria sarà il primo ad uscirne, e quindi possiamo pensare che il primo campione ad essere letto sarà spinto sempre più verso l'uscita dai campioni introdotti in seguito. La velocità di scrittura di memorie di questo tipo deve essere idonea alla velocità del convertitore A/D: non devono comportarsi come un collo di bottiglia rispetto alla velocità del trasduttore, bensì essere quantomeno in grado di memorizzare istantaneamente i dati da esso inviati, in modo da non subire perdite di elementi numerici. Le dimensioni della memoria variano da alcuni kB fino ad alcune centinaia di kB, a seconda del numero di campioni che il sistema è in grado di immagazzinare

Blocco di Trigger

Il blocco di trigger, come già accennato, serve a gestire essenzialmente le modalità di lettura di dati dalla memoria; in questo modo si riesce a posizionare la finetra temporale di osservazione in prossimità di un dato istante da noi selezionato (o in prossimità di un determinato evento); le modalità di trigger possono essere, aldilà degli aspetti quali pre-trigger e post-trigger, molto simili a quelle dell'oscilloscopio analogico: incontro di un certo fronte, ad un certo livello di tensione, con una certa pendenza. Oltre a queste condizioni, il DSO permette di introdurre alcune condizioni logiche, in grado di far scattare la commutazione del circuito di trigger.

Sistema di rappresentazione

Come tutti i blocchi finora analizzati sono stati migliorati rispetto al vecchio oscilloscopio digitale, così è stato anche per il sistema di rappresentazione del segnale su schermo: la tecnologia di base per un certo tempo è sempre stata la CRT (ormai sempre più surclassata dalla LCD), però basata su di una scansione tipo raster, dettata da una deflessione magnetica. Grazie a questa tecnologia si è riusciti a realizzare schermi più grandi, potendo utilizzare deflessioni maggiori, e anche ad introdurre, in questi schermi, un maggior numero di informazioni, tra cui i colori (in grado di meglio distinguere diverse caratteristiche dei segnali).

5.2 Modalità di campionamento

Il campionamento è una delle procedure più importanti per il funzionamento complessivo dell'oscilloscopio numerico; analizziamone le due principali categorie.

5.2.1 Campionamento in real time

Quando abbiamo un segnale non ripetitivo, quale ad esempio un transitorio, dobbiamo concetrare tutta l'acquisizione sull'intervallo di tempo che ci interessa studiare del fenomeno, in modo da avere campioni sufficienti da permetterci una rappresentazione soddisfacente. Il clock dell'oscilloscopio gestisce tutte le operazioni di acquisizione dati; poichè all'ingresso dell'oscilloscopio abbiamo un filtro anti-aliasing, ossia un passa basso, sarà sufficiente avere, ai fini di soddisfare il criterio di Nyquist, una frequenza superiore al doppio della banda assoluta di questo filtro in ingresso. Può capitare di voler campionare addirittura a 4 volte la banda assoluta del filtro, poichè, se il criterio di Nyquist è rispettato con condizioni minimali ($F_S = 2B_{abs}$), si ha una riduzione dell'ampiezza di circa 3 dB (ossia il 30 %, dal momento che parliamo di decibel in contesto di grandezze lineari).

5.2.2 Campionamento di segnali ripetitivi

Caso più frequente e che svilupperemo meglio riguarda i segnali ripetitivi, ossia segnali ciclici, che continuano a riproporsi in maniera uguale, nel tempo. Caso particolare di questi segnali sono i periodici.

I segnali ripetitivi sono interessanti in quanto è possibile non rispettare con essi il criterio di Nyquist (o meglio, non rispettarlo in modo furbo): anzichè campionare in un singolo periodo del segnale tutti i campioni necessari alla ricostruzione del segnale finale, è possibile, mediante due sostanziali tecniche, sottocampionare anche fortemente, ed ottenere in uscita un segnale molto fedele a quello reale. Al fine di effettuare un sottocampionamento, ovviamente, è necessaria la stazionarietà del segnale, ossia la garanzia di non percepire variazioni nel tempo del segnale. Come già accennato, a seconda del tipo di metodo utilizzato per leggere i campioni in stato di sottocampionamento, possiamo distinguere due tecniche:

- Campionamento sequenziale;
- Campionamento casuale.

5.2.3 Campionamento sequenziale

Nel caso di campionamento sequenziale, consideriamo la seguente idea: dal momento che sottocampioniamo, e che il segnale è ripetitivo, per ogni periodo campioniamo un singolo elemento del segnale; dato un certo ritardo Δt rispetto all'inizio del segnale, e un numero k intero, semplicemente, per ogni k-esimo periodo, campioneremo il k-esimo valore di Δt: al primo scorrimento del segnale campioneremo Δt, al secondo scorrimento $2\Delta t$, al terzo $3\Delta t$, e così via fino a terminare.

La frequenza scelta, come:

$$f_{eq} = \frac{1}{\Delta t}$$

Deve essere sufficientemente elevata da poter soddisfare il criterio di Nyquist. Questo metodo presenta alcuni inconvenienti: supponiamo di disporre di un campionatore a 100 MHz, e di un segnale da 1 GHz: potremo campionare un segnale utile ogni dieci ripetizioni del segnale, e quindi il processo sarà molto lento; inoltre, si ha una forte dipendenza dall'evento di trigger del campionamento: non è possibile acquisire porzioni di segnali precedenti all'evento di trigger. Il fatto di disporre di un processo lento, comunque, ci permette di poter utilizzare convertitori A/D lenti, accurati, e meno costosi.

Questo metodo viene utilizzato per strumenti analizzatori di particolari forme d'onda, a frequenze estremamente elevate, ottenendo una buona accuratezza nella conversione A/D.

5.2.4 Campionamento casuale

Si tratta di una tecnica che ora citeremo, ma che normalmente non viene utilizzata nell'oscilloscopio digitale: essa è basata sul raccogliere campioni, anzichè in maniera ordinata e sequenziale come nel precedente esempio, in maniera del tutto casuale. Questo metodo è il più utilizzato, generalmente, negli oscilloscopi digitali. Si campionano dunque segnali in tempi casuali, però sempre alla stessa frequenza di lavoro, f_s; una volta raccolti, i campioni vengono riordinati in memoria in ordine temporale crescente, rispetto al tempo di trigger, istante di riferimento. La frequenza f_s utilizzata per il processo di campionamento non è correlata ai campioni raccolti, dunque vi è una buona probabilità di non effettuare ripetizioni, ossia di non raccogliere due volte il medesimo campione. Il trigger ed il generatore di impulsi sono desincronizzati: se così non fosse, si avrebbe sempre lo stesso istante campionato.

Consideriamo PV una porzione di segnale, che intendiamo presentare con il nostro DSO; abbiamo N campioni raccolti dal campionatore, in grado di presentarlo. Ognuno di questi campioni, dunque, serve a rappresentare un certo time slot, ossia un certo insieme di valori. Quello che intendiamo fare è, considerando l'intero PV separato in N time slot, ricollocare ogni intervallo temporale nel proprio time slot (mediante un algoritmo di sort). La durata di ogni time slot, T_S, sarà dunque:

$$T_S = \frac{PV}{N}$$

Considerando la distanza di ogni campione con l'evento di trigger, stabilito dall'operatore, sarà possibile inserire ogni campione nel proprio time slot, ordinandoli dunque cronologicamente come desideriamo.

La porzione visualizzata, PV, si sceglie con la selezione base tempi (Time/Div); N rappresenta il numero di punti campionati, o meglio di punti presenti sullo schermo. La durata del time slot T_S, definita prima, dipende dal coefficiente di deflessione verticale che scegliamo con l'apposita manopola. Un time slot, in un DSO, ha una durata minima, dipendente dalla risoluzione temporale con la quale si può misurare la distanza temporale ΔT tra l'evento di trigger ed il campione che ci interessa. Poichè la frequenza equivalente, f_s, vale:

$$f_s = \frac{1}{\Delta T}$$

Allora sostanzialmente, ciò che limita la banda dell'oscilloscopio, è il filtro anti-aliasing introdotto sull'ingresso.

Convertitori A/D

Al fine di campionare velocemente, come già detto nelle introduzioni, è necessario avere un frontend in grado di convertire velocemente da segnale analogico a segnale digitale. Ciò che si utilizza, è un convertitore basato sul metodo delle approssimazioni successive:

Si ha in sostanza un comparatore di soglia, con un sample/hold, che confronta una tensione (analogica) con l'uscita di un covertitore D/A, ossia un dispositivo duale all'A/D: dato un numero in ingresso, fa uscire una tensione. Il numero in ingresso al D/A, e quindi la tensione che andrà continuamente confrontata nel comparatore di soglia, deriverà da un circuito digitale con logica dicotomica: ad ogni step, a seconda del risultato del comparatore di soglia, la logica dicotomica aggiunge o toglie (a seconda del fatto che stiamo sbagliando in difetto o in eccesso) metà della quantità, fino ad avvicinarci al risultato finale. La logica più semplice utilizzata è quella a 3 bit, che introduce una quantizzazione a 8 livelli.

Il fatto di lavorare con una quantizzazione così bassa, dovuto alle esigenze di velocità, ci costringono a non pretendere molto in fatto di risoluzione.

Se le frequenze di campionamento f_s sono addirittura più alte di quelle introdotte, i circuiti finora utilizzati non sono sufficienti, poichè si dovrà ricorrere a convertitori FLASH, in cui T_c è pari al tempo di clock (dal momento che questi convertitori sono parallelo e molto rapidi).

In sostanza, abbiamo un certo numero di comparatori di soglia con la stessa V_x di riferimento in uno dei morsetti di ingresso, e la tensione comparata sull'altro morsetto. Un certo numero di comparatori commuterà, altri no, e così la V_x incognita verrà determinata dal numero di comparatori che ha commutato da una parte o dall'altra, scelta dal circuito logico cui son collegate tutte le uscite (che in sostanza è un multiplexer).

Non disponendo di tecnologie di questo tipo, esiste un metodo alternativo: utilizzare molti convertitori lenti anche se le f_c sono elevate. Usando diversi campionatori sample/hold, in parallelo tra loro, si riesce ad ottenere una velocità di conversione n volte superiore a quella di un singolo campionatore, dove n è il numero di dispositivi utilizzati in parallelo tra loro. Il multiplexer di nuovo potrà smistare i segnali in arrivo dai vari S/H.

5.3 Modalità di trigger

Come già detto, lo scopo dei trigger, nei DSO, è leggermente diverso rispetto a quello che avevano in ambito di oscilloscopi analogici; potremmo definire infatti come trigger l'evento attorno alq uale si raccolgono e presentano i campioni della finestra che intendiamo visualizzare. Esso rappresenta di fatto un riferimento temporale, per gli N campioni conservati in memoria in seguito all'operazione di campionamento, parlando di trigger real time; parlando di trigger riguardante segnali ripetitivi, esso determina un punto di riferimento a partire dal quale individuare una ripetizione del segnale da analizzare.

A differenza che nell'oscilloscopio analogico, nell'oscilloscopio digitale, oltre a tutti i comandi classici già introdotti nell'oscilloscopio analogico, si dispone di alcune possibilità extra, quali il pre-trigger ed il post-trigger; inoltre, l'enorme versatilità dello strumento permette di effettuare trigger anche a partire da più di un segnale contemporaneamente, o utilizzando particolari condizioni logiche.

La versatilità consiste nella possibilità di contare i campioni, fissando l'evento di trigger in modo che, dopo di esso, si abbiano M ulteriori campionamenti; considerando K il numero di campioni memorizzabili, la larghezza della memoria. Il parametro inseribile dall'utente è dunque proprio questo M: possiamo scegliere quanti campioni devono ancora essere memorizzati, dopo il nostro punto di riferimento, ossia l'istante di trigger.

Trigger in Real Time

Nell'ambito del real time, per esempio durante l'acquisizione di un segnale transitorio, è necessario memorizzare subito un certo numero di campioni, e stoppare il tutto; regolando il parametro M in modo da poter posizionare il trigger in una posizione piuttosto che un'altra, potremo visualizzare sullo schermo l'informazione che più ci interesserà. Facciamo alcuni esempi:

- Se $M = 0$, si ha che nessun campione viene memorizzato dopo l'istante di trigger, e dunque possiamo dire di avere esclusivamente presentato segnali in pre-trigger;
- Se $M = K$ vuol dire che abbiamo campionato tutti e soli i tempi dopo l'istante di trigger, e dunque visualizzeremo qualcosa di simile a ciò che si visualizzerebbe nell'oscilloscopio analogico (si tenga conto che però ora siamo in grado di visualizzare fenomeni transitori).
- Se $0 < M < K$, la memoria contiene una porzione di segnale a cavallo dell'evento di trigger: si ha sia pre-trigger che post-trigger
- Se $M > K$, si visualizza una porzione di segnale ritardata rispetto all'evento di trigger

Trigger con segnali ripetitivi

Poco c'è da aggiungere rispetto a prima: sappiamo che, nel caso di campionamento di eventi casuali, il trigger rappresenta un punto temporale di riferimento rispetto cui effettuare il sort dei campioni; sono assolutamente valide tutte le opzioni finora presentate per quanto riguarda il trigger in real time.

5.4 Presentazione

Elemento finora trascurato nella trattazione degli oscilloscopi numerici è la presentazione del risultato: un oscilloscopio digitale deve essere in grado, oltre che di effettuare le varie operazioni di immagazzinamento ed elaborazione dei dati, di presentare un risultato sullo schermo. La presentazione, oltre alla forma d'onda, ossia oltre all'andamento temporale (e non solo) della porzione di segnale di nostro interesse, deve essere in grado di presentare indicazioni alfanumeriche riguardanti il segnale (quali sensibilità verticale ed orizzontale, modalità di trigger impiegate per la presentazione, ed anche misure di vario genere). Si noti che, dal momento che la curva presentata è di fatto una ricostruzione per punti, la qualità della traccia è generalmente inferiore a quella presentata da un oscilloscopio analogico, che riesce a dare più la sensazione della continuità nel disegno.

Esistono sostanzialmente due metodi di deflessione (parlando di schermi a CRT):

- Deflessione vettoriale: mediante un campo elettrico, le placchette riescono a deflettere sulle componenti X e Y di un piano cartesiano gli elettroni, che vengono sparati sullo schermo. Questo è il metodo che classicamente si utilizza negli oscilloscopi analogici: esso permette di definire in maniera migliore la traccia, ma non permette di utilizzare colori (cosa che spesso può tornare utile).

- Deflessione raster: mediante il campo magnetico, si effettua una scansione di riga e di quadro. Lo schermo CRT viene sostanzialmente considerato come una matrice di pixel di dimensioni HxL. I dati vengono memorizzati in una memoria a matrice, detta memoria di schermo, nella quale vengono mappati. Ciascuna delle parole contenute per elemento della matrice contiene colore ed intensità; al momento della scansione, effettuata ciclicamente (in modo da far persistere l'immagine sullo schermo), vengono impresse sullo schermo le caselle per ciascuno dei pixel. Questa tecnica permette una peggior qualità della traccia, ma al contempo permette i colori, e schermi di dimensioni più grandi (entrambe le cose interessanti). Questo tipo di deflessione è la più utilizzata, nell'ambito dei DSO (come anche per i calcolatori elettronici).

5.4.1 Tecniche di ricostruzione

Il segnale che intendiamo rappresentare sullo schermo dell'oscilloscopio digitale è di durata limitata; dalla Teoria dei Segnali, sappiamo che, a durata finita nel tempo, equivale uno spettro a banda di larghezza infinita. Dal momento che il nostro obiettivo è quello di presentare sullo schermo una traccia più simile possibile a quella che dovrebbe essere nella realtà, partendo da un numero limitato di campioni, dovremo effettuare alcuni accorgimenti: in teoria, avremmo bisogno di infiniti campioni, al fine di rappresentare in maniera corretta il segnale; nella pratica, ciò non è assolutamente possibile, come possiamo immaginare.

Il nostro obiettivo primario è eliminare l'aliasing: dobbiamo evitare che vi sia una sovrapposizione di code, generanti interferenze e quindi perdita di informazioni riguardanti il segnale da rappresentare. Come prima cosa, dunque, dobbiamo fare non solo in modo che la frequenza di campionamento rispetti il criterio di Nyquist, ma che lo sovrasoddisfi, in modo da avere dunque:

$$F_S \gg 2B$$

Dove B è la banda del segnale in questione.

Poichè non è generalmente possibile presentare una forma d'onda in modo ottimale, ossia fornendo una parvenza di continuità alla traccia presentata, sarà necessario ricorrere ad un processo di interpolazione, aggiungendo campioni negli intervalli di tempo in cui mancano quelli che non abbiamo potuto acquisire per limiti tecnici. Esistono metodi di interpolazione lineare, ossia consistenti nell'aggiungere semplicemente segmenti congiungenti i vari punti adiacenti tra loro; possiamo immaginare che, se abbiamo a disposizione pochi punti, questo metodo sia molto poco soddisfacente.

Altra tecnica di interpolazione è basata sull'utilizzo della funzione $\mathrm{sinc}(x)$:

$$\mathrm{sinc}(x) = \frac{\sin(x)}{x}$$

Nella fatispecie, dovremmo utilizzare una funzione di questo genere:

$$x(t) = \sum_{n=-\infty}^{+\infty} x(nT_S) \frac{\sin(\pi F_S(t - nT_S)}{\pi F_S(t - nT_S)}$$

Dove si definisce la frequenza di campionamento F_S come:

$$F_S = \frac{1}{T_S}$$

Si noti che, come già detto in precedenza, F_S potrebbe essere anche molto al di sotto della frequenza soddisfante il criterio di Nyquist, in caso di rappresentazione di segnali ripetitivi; ciò che abbiamo detto prima, è utile al fine di rappresentazioni di segnali qualunque.

Nel caso di segnali ripetitivi, si può utilizzare una sorta di frequenza equivalente, F_{eq}, definita come:

$$F_{eq} = \frac{1}{T_{eq}} = \frac{1}{\frac{PV}{N_V}}$$

Dove PV è la porzione di schermo presentata, come già avevamo detto in precedenza; N_V rappresenta invece il numero di punti della traccia presenti sullo schermo: esso è un parametro prettamente legato alle caratteristiche dello schermo, e dovrebbe essere dichiarato dal costruttore. Al contrario, PV dipende da una scelta dell'utente, ossia dalla scelta della base tempi dell'operatore.

L'oscilloscopio digitale non è semplice da utilizzare: l'operatore alle misure deve essere in grado di capire come utilizzarlo, e capire se esso presenta risultati evidentemente sbagliati: se variando ad esempio la taratura dell'asse dei tempi (mediante la solita manopola Time/Div) il DSO varia notevolmente i parametri della forma d'onda, vuol dire che qualcosa nel processo di rappresentazione è andato storto. Sarà necessario valutare parametri più idonei del Time/Div, al fine di rappresentare in maniera corretta i segnali sullo schermo;

si noti che tutti questi accorgimenti devono essere effettuati dall'operatore, che deve acquisire una certa manualità nell'effettuare operazioni di questo genere.

5.4.2 Prestazioni di un DSO

Altra cosa che l'operatore deve essere in grado di conoscere è la dichiarazione delle prestazioni: le prestazioni, dichiarate dal costruttore sui vari datasheet e manuali, devono essere interpretate in maniera corretta dall'utente del DSO. Proponiamo alcuni esempi, sui quali bisognerebbe fare molta attenzione.

Banda passante del DSO

Molto spesso, parlando di banda passante, si fraintende, pensando ad un elemento sbagliato: la banda passante non rappresenta ciò che l'oscilloscopio è in grado di rappresentare in maniera più o meno corretta, bensì la frequenza intorno cui lo strumento inizia a distorcere il segnale in maniera violenta. Solitamente, parlando di real time, la banda passante dichiarata da molti costruttori è circa uguale a:

$$B_{RT} \simeq \frac{1}{4} F_{S,max}$$

Dove $F_{S,max}$ è la massima frequenza di campionamento del dispositivo. Parlando di real time, il vero collo di bottiglia dell'oscilloscopio non è il fatto che vi siano effetti di filtraggio dovuti ad eventuali non idealità dei circuiti, bensì la velocità del campionatore.

Per quanto riguarda i segnali ripetitivi, sappiamo che è possibile sottocampionare fortemente, dunque il discorso effettivamente è diverso: è possibile leggere una banda passante del dispositivo molto più elevata rispetto a quella di campionamento, in quanto si parla delle frequenze che effettivamente provocano problemi a livello di filtraggio. Possiamo dunque dire che:

$$B_{RIP} \gg F_{S,max}$$

Bisogna essere in grado di saper distinguere questi due casi, in modo di capire precisamente di cosa si parla.

Risoluzione ed accuratezza verticale

Il rapporto segnale/rumore, ossia l'indice della risoluzione teorica della deflessione verticale, dovrebbe dipendere dal convertitore A/D; quello che capita è che, aumentando i bit del quantizzatore, si riesce ad aumentare il rapporto SNR, e così la risoluzione verticale del sistema; si sappia però che, aumentando i bit del quantizzatore (e dunque i livelli di quantizzazione), si tende a degenerare prima verso la soglia del $6n|_{dB}$, ossia il limite asintotico (valutato in un diagramma in decibel) cui tende il rapporto segnale/rumore.

L'accuratezza legata alla sola componente verticale, viene solitamente fornita in termini percentuali rispetto al fondo scala della portata selezionata sull'asse delle ordinate, e dovrebbe essere compresa tra $0,2\% \div 2\%$. Talvolta può capitar di avere a che fare con un formato diverso, considerando ancora un contributo assegnato solo al fondo scala, più una quantità fissa (il tutto in percentuale)

Accuratezza orizzontale

L'accuratezza orizzontale, ossia l'accuratezza dell'asse dei tempi, è determinata da due principali elementi:

- L'accuratezza del clock che determina gli istanti di campionamento (solitamente, accuratezza molto buona, visto che si ha a che fare con dispositivi piuttosto precisi, come già visto);
- L'accuratezza con cui si misurano i ritardi dei campioni rispetto all'evento di trigger (come abbiamo visto, parlando di campionamento casuale).

Di solito, si presenta l'incertezza in maniera analoga ai voltmetri digitali: percentuale del valore letto, più percentuale dovuta alla sensibilità orizzontale scelta (e quindi a seconda della portata utilizzata per riscalare il dominio dei tempi visualizzabile sullo schermo del dispositivo).

Capitolo 6

Ponte di Wheatstone

Si parla spesso di misura di resistenze mediante tester, o altri dispositivi, spesso basati sul metodo voltamperometrico: misurando contemporaneamente corrente e tensione in un determinato punto, utilizzando la legge di Ohm, si può misurare la resistenza nel punto come:

$$R = \frac{V}{i}$$

Esistono metodi alternativi per la misura di una resistenza, basati sull'utilizzo di circuiti a ponte, alimentati da un generatore di tensione, o anche di corrente, considerato ideale per ipotesi. Nella fatispecie, noi considereremo per ipotesi di alimentare il ponte mediante un generatore di tensione ideale, E.

Il ponte è costituito da due resistori noti, R_a e R_b, un resistore campione R_c, e il resistore incognito, che è nostra intenzione misurare, ossia R_x.

Si tratta di un mezzo in realtà assolutamente obsoleto negli anni in cui ci troviamo, tuttavia esso è molto interessante da studiare in quanto ha posto le basi per le misure di impedenza, ossia di elementi contenenti, oltre ad una parte resistiva, una parte reattiva, immaginaria. Il ramo centrale del ponte è collegato ad un rilevatore di tensione, un voltmetro, che consideremo per ipotesi ideale, ossia dotato di impedenza di ingresso molto elevata. Consideriamo come tensione di uscita V_u la tensione ai capi del voltmetro, normalizzata rispetto all'ingresso E.

Incominciamo a studiare il funzionamento teorico del ponte, calcolando la sua funzione trasferimento, ossia il rapporto tra la tensione di uscita V_u e l'ingresso E. Mediante le nozioni di Elettrotecnica:

$$\frac{V_u}{E} = \frac{R_a R_c}{R_a + R_b} \frac{1 - \frac{R_b}{R_a R_c} R_c}{R_x + R_c}$$

Mediante passaggi algebrici, troviamo:

$$\frac{R_b}{R_a + R_b} \frac{1 - \frac{R_b}{R_a R_c} R_c}{\frac{R_b}{R_c R_a} R_x + \frac{R_b}{R_a}}$$

Consideriamo ora per comodità un cambio di variabile, al fine di rendere più leggibile la funzione di trasferimento:

$$X = \frac{R_b}{R_a R_c} R_x$$

La funzione di trasferimento diventerà:

$$\frac{V_u}{E} = \frac{R_b}{R_a + R_b} \frac{1 - X}{X + \frac{R_b}{R_a}}$$

A questo punto possiamo intuitivamente capire come è possibile utilizzare questo ponte per effettuare misurazioni: se analizziamo la funzione $H(X)$, ossia la funzione di trasferimento del circuito al variare però della variabile X prima definita, vediamo che:

$$H(X) = 0 \Longrightarrow (1 - X) = 0 \Longrightarrow X = 1$$

Nel punto $X = 1$, $V_u = 0$, e quindi si ha uno zero di trasmissione (ossia uno zero della funzione di trasferimento, anche se consideriamo la variazione non rispetto alle frequenze come convenzionalmente si fa, ma rispetto alla variabile X).

6.1 Ponte di Wheatstone all'equilibrio

Solitamente, si sceglie utilizzare il ponte di Wheatstone come misuratore di resistenze mediante metodo del confronto, considerandolo all'equilibrio, ossia sfruttando lo zero di trasmissione appena ricavato: se infatti abbiamo il ponte all'equilibrio, sappiamo per certo che $X = 1$, ma quindi, sostituendo a X il suo valore iniziale, vediamo che:

$$X = 1 \Longrightarrow \frac{R_b}{R_a R_c} R_x = 1 \Longrightarrow R_x = \frac{R_a R_c}{R_b}$$

Per utilizzare a questo punto nella pratica il ponte in maniera intelligente, mediante il metodo del confronto, dobbiamo avere due resistori stabili, con valori fissi, mentre uno a valori variabili, in modo da poterne variare il valore fino a ottenere il fatidico valore di equilibrio, ossia lo zero di trasmissione: $V_u = 0$. A queste condizioni, sappiamo per certo che la relazione $X = 1$ è utilizzabile, e quindi mediante la formula prima esposta potremo calcolare il valore del resistore ignoto R_x, a partire dalla misurazione di $V_u \to 0$, e data una buona definizione del resistore campione (potremmo per esempio considerare R_a.

Quello qua esposto è l'esempio più semplice di utilizzo: è ovviamente possibile utilizzare anche più di un resistore campione, tuttavia si tenga presente che un resistore variabile si può considerare campione se e solo se si ha nota la curva di taratura del medesimo, ottenuta dal costruttore, o confrontando il resistore con altri di qualità superiore ad esso.

6.1.1 Risoluzione del Ponte in caso di voltmetro ideale

Ci sono due filosofie di pensiero, sulla variabilità dei resistori: è possibile considerare resistori variabili con continuità, o resistori che variano con valori discreti.

Si ha di fatto, con resistori variabili con continuità, una potenziale risoluzione infinita del fattore: provando a pensare in termini di Analisi Matematica, se si ha la possibilità di variare sulla retta reale non bucata, si ha di fatto la possibilità di incontrare numeri con infinite cifre decimali (si noti che stiamo parlando di risoluzione, non di accuratezza); dato un voltmetro molto vicino all'idealità, ed un sistema privo di rumori di qualsiasi tipo, possiamo sperare che un certo numero di queste cifre decimali possa essere significativo.

Se si utilizzano campioni con variazione discreta, la risoluzione sarà certamente legata alla minima variazione di resistenza, al peso della resistenza più piccola nell'ambito del sistema. Considerando dunque R_a il resistore campione, la risoluzione, ossia ΔX, la minima variazione della grandezza che abbiamo finora studiato, sarà:

$$\Delta X = \frac{R_x}{\Delta R_a R_c} R_b$$

In questo ambito, poteva essere una scelta più felice come resistenza variabile la R_b: essendo essa al numeratore, è più semplice da trattare a livello di calcoli. Presentiamo una soluzione alternativa al problema del calcolo della risoluzione, a partire quindi da R_b come resistore variabile, con scatti discreti.

$$\Delta X = \frac{R_x}{R_a R_c} \Delta R_b$$

Date tutte queste ipotesi, possiamo dire che quindi una variazione del resistore R_x sia direttamente implicabile ad una variazione del resistore R_b, e dunque possiamo dire che, considerando un certo valore di resistenza $R_{x,0}$ associato dunque ad un certo valore $R_{b,0}$, tale per cui:

$$R_{x,0} = \frac{R_c R_a}{R_{b,0}}$$

Valga la seguente relazione tra le variazioni di R_x e R_b:

$$\left| \frac{\Delta R_x}{R_{x,0}} \right| = \left| \frac{\Delta R_b}{R_{b,0}} \right|$$

Tutto ciò è ovviamente vero solo nel contesto di tutte le ipotesi finora enunciate, nella fatispecie, in un intorno di $X = 0$.

Considerando dunque una discretizzazione nella variazione dei resistori, c'è da tener conto di un fattore importante: dal momento che il valore del resistore campione non varia con continuità, non sarà possibile in generale ottenere, dalle misurazioni, il valore di R_x (e quindi del resistore campione R_b) tale per cui $V_u = 0$; ciò che possiamo fare ad esempio è scegliere come valore di R_x quello che determina la tensione V_u più vicina a 0, tra tutti i campioni disponibili. Supponendo ad esempio di aver a disposizione due campioni X_1 e X_2 (per semplificare l'esempio: in realtà di solito si dispone di un numero molto più elevato di soli due campioni): il valore della resistenza misuranda, R_x, si può approssimare con il valore della resistenza che più avvicina il ponte ad una condizione di equilibrio. Supponendo che questo valore sia nel nostro esempio X_2:

$$R_x \simeq R_{x,2} = \frac{R_a R_c}{R_{b,2}}$$

Ricollegandoci dunque all'espressione precedentemente vista per quanto riguarda i campioni discreti, possiamo dire che l'ampiezza della fascia di discretizzazione, ossia la risoluzione causata dall'utilizzo di campioni R_b variabili con soli valori discreti, dipenda, in termini relativi, come:

$$\frac{|\Delta R_x|}{R_{x,2}} = \frac{|\Delta R_b|}{R_{b,2}}$$

A partire da quest'espressione, possiamo considerare la risoluzione assoluta come semiampiezza della fascia di valori ΔR_x:

$$\delta_{ass} = \pm \frac{\Delta R_x}{2}$$

Anzichè scegliere il valore di R_x semplicemente considerando quello che porta il ponte ad una situazione più vicina all'equilibrio, è possibile adottare mezzi statistici, quali l'interpolazione: considerando di nuovo due punti X_1 e X_2, con le relative tensioni $V_{u,1}$ e $V_{u,2}$ in uscita del ponte considerando gli stati appunto X_2 e X_2, il valore di R_x si può calcolare mediante metodi geometrici, utilizzando semplicemente la formula della retta passante per due punti: considerando come due punti gli stati $(X_1; \frac{V_{u,1}}{E})$ e $(X_2; \frac{V_{u,2}}{E})$, utilizzando la formula della retta passante per due punti, non considerando il termine E che tanto si andrà a semplificare, si ricava:

$$\frac{V_y + V_{u,2}}{V_{u,1} + V_{u,2}} = \frac{R_x - X_2}{X_1 - X_2}$$

Poichè il nostro obiettivo è avere il ponte all'equilibrio, la tensione y dovrà essere posta uguale a 0, al fine di ottenere il punto che ci interessa della retta interpolante in questione:

$$\frac{0 + V_{u,2}}{V_{u,1} + V_{u,2}} = \frac{R_x - X_2}{X_1 - X_2} \Longrightarrow \frac{V_{u,2}(X_1 - X_2)}{V_{u,1} + V_{u,2}} = R_x - X_2$$

$$R_x = \frac{X_1 V_{u,2} + X_2 V_{u,1}}{V_{u,1} + V_{u,2}}$$

Dal momento che il rilevatore dello stato del ponte, ossia il voltmetro, è ideale, possiamo dire che la risoluzione sia unicamente limitata dalla risoluzione del campione, ossia di R_b: abbiamo già detto come è possibile ricavare R_x e ΔR_x in funzione di R_b, e dunque, possiamo dire che questa, in un sistema con voltmetro ideale, sia l'unica fonte di riduzione della risoluzione. Possiamo dunque scrivere che:

$$\frac{\Delta R_x}{X_2 - X_1} = \frac{\Delta V_u}{V_{u,1} + V_{u,2}}$$

Da qua si può dunque facilmente ricavare:

$$\Delta R_x = \frac{X_2 - X_1}{V_{u,1} + V_{u,2}}$$

Pertanto abbiamo quantificato l'apporto del solo resistore campione R_b sulla risoluzione del sistema.

6.1.2 Risoluzione del Ponte in caso di voltmetro non ideale

Come possiamo immaginare un voltmetro ideale, o comunque un sistema determinante lo stato del ponte senza incertezza, non esiste nel nostro mondo fisico, quindi sarà necessario tener conto degli effetti di non idealità introdotti dal voltmetro. A causa di questi effetti di non idealità, la lettura del valore $V_u = 0$ è limitata dalla presenza di rumori di diverso tipo. Considerando dunque l'elemento variabile con continuità, non avremo valori precisi con i quali approssimare R_x, ma una fascia di valori attorno alla linea ideale, ossia alla linea che prima di considerare effetti dispersivi del voltmetro approssimava il comportamento del resistore variabile.

Possiamo però ragionevolmente stimare il valore di R_x come il centro della fascia, la cui ampiezza si può determinare a partire dalla conoscenza della retta osculatrice della caratteristica di trasferimento, del ponte, in un intorno di $V_u = 0$.

Conoscendo dunque l'andamento della funzione $V_u(R_x)$, è possibile determinare la pendenza di questa retta come:

$$p = \left.\frac{\partial V_u}{\partial R_x}\right|_{V_u=0}$$

A questo punto, è possibile calcolare la risoluzione assoluta (sempre interpretabile come semiampiezza della fascia di incertezza) come:

$$|\Delta R_x| = \left|\frac{\Delta V_{u,0}}{p}\right|$$

Dove il $\Delta V_{u,0}$ è la risoluzione del ponte in un intorno di tensione di uscita $V_u \simeq 0$.

Osservando il circuito del ponte, considerandolo in un intorno del punto di equilibrio (e quindi tensione circa nulla sul ramo del voltmetro), possiamo considerare questa $V_u \simeq 0$, quantificandola come differenza delle tensioni della parte destra del ponte $(R_x + R_c)$ e della parte sinistra $(R_a - R_b)$, considerando dunque positiva la tensione da sinistra verso destra:

$$V_u(R_x) = E\left(\frac{R_c}{R_x + R_c} - \frac{R_b}{R_a + R_b}\right) = E\left(\frac{R_c(R_a + R_b) - R_b(R_x + R_c)}{(R_a + R_b)(R_x + R_c)}\right) = E\frac{R_cR_a - R_bR_c}{(R_a + R_b)(R_x + R_c)}$$

Questa è l'espressione della quale calcoleremo la derivata parziale, al fine di ottenere il coefficiente angolare della retta di cui parlavamo prima:

$$\left.\frac{\partial V_u}{\partial R_x}\right|_{V_u=0} = -E\frac{R_b^2}{R_c(R_a + R_b)^2}$$

Utilizzando infine la formula di prima:

$$|\Delta R_x| = \left|\frac{\Delta V_{u,0}}{p}\right| \Longrightarrow |\Delta R_x| = \frac{R_c(R_a + R_b)^2}{ER_b^2}\,|\Delta V_u|$$

Calcolare questo ΔV_u non è banale: per farlo, è possibile sfruttare le nostre conoscenze sulla variazione del resistore campione ΔR_b, e così utilizzare le formule precedentemente viste.

Talvolta può capitare di non disporre di un voltmetro tarato, ossia che non presenta una scala di lettura. In tal caso, dato il resistore campione R_b con un determinato valore $R_{b,0}$, è possibile stimare la variazione di tensione, σ_x, mediante lo studio del ponte sbilanciato:

$$\sigma_x = \frac{\frac{\Delta R_b}{R_{b,0}}}{\frac{\Delta V_u}{\delta V_u}}$$

Dove δV_u è la minima variazione apprezzabile sul rilevatore.

Riassumendo, abbiamo terminato lo studio della risoluzione del ponte di Wheatstone, in caso ideale (con variazione discreta dell'elemento campione), r_1, ed in caso reale, r_2, con variazione continua dell'elemento campione (ma impedenza del voltmetro non infinita e quindi effetti di consumo). I due contributi saranno:

$$r_1 = \frac{1}{2}\frac{|\Delta R_x|}{R_x} = \frac{|\Delta R_b|}{R_b}$$

$$r_2 = \frac{|\Delta R_x|}{R_x} = \frac{1}{R_x \cdot E}\frac{R_c(R_a + R_b)^2}{R_b^2}|\Delta V_u| = \frac{1}{E}\frac{(R_a + R_b)^2}{R_a R_b}|\Delta V_u|$$

Abbiamo trattato questi due casi separatamente per comodità: in realtà, di solito, si ha a che fare con un voltmetro non ideale ed un resistore campione variabile discretamente; i due effetti si combineranno dunque linearmente, mediante una somma:

$$r_{eq} = r_1 + r_2$$

Esistono alcuni metodi per migliorare la risoluzione (si noti che non stiamo ancora parlando di accuratezza) del ponte di Wheatstone: un metodo banale è migliorare la risoluzione del campione variabile, ossia considerare un campione che varia discretamente, ma con passi più vicini tra loro. Altro metodo scontato è migliorare la sensibilità del rilevatore, ossia del voltmetro, in un intorno di $V_{mis} = 0$: non è detto che il diagramma di taratura del voltmetro sia infatti lineare: se si ha una risoluzione migliore in prossimità dello 0, allora la risoluzione complessiva del sistema non potrà che riceverne benefici. Un'arma a doppio taglio che può migliorare la risoluzione del ponte è l'aumento della tensione di alimentazione del medesimo, E: ciò migliora sì la risoluzione, poichè r_2 è inversamente proporzionale ad E, ma aumentando la polarizzazione si aumenta di sicuro anche l'effetto di consumo, rischiando di andare a perdere in accuratezza, come vedremo in seguito, dissipando maggiore potenza.

Accuratezza del Ponte di Wheatstone

Sicuramente il primo contributo di incertezza nella misura del ponte sarà dato dalle incertezze sulle resistenze che lo compongono: poichè la misura indiretta del resistore R_x si ricava mediante una relazione di prodotti e quozienti:

$$R_x = \frac{R_a R_c}{R_b}$$

Dovremo calcolare l'incertezza relativa equivalente ε_x mediante somma di tutte le incertezze relative su di ogni singolo resistore:

$$\varepsilon_x = \varepsilon_a + \varepsilon_b + \varepsilon_c$$

Consideriamo il caso reale, più interessante: possiamo dire che, in esso, interverrà anche il termine di risoluzione equivalente prima introdotto, r_{eq}, e quindi:

$$R_x = \frac{R_a R_c}{R_b} \pm r_{eq} \Longrightarrow \delta R_x = \delta\left(\frac{R_a R_c}{R_b}\right) + r_{eq}$$

Dividendo tutto per R_x, si ottiene:

$$\frac{\delta R_x}{R_x} = \varepsilon_a + \varepsilon_b + \varepsilon_c + \frac{\delta r_{eq}}{R_x}$$

Si noti che, in un sistema ben progettato, l'ultimo termine non dovrebbe comparire: in questo modo potremmo, in prima approssimazione, considerare il nostro ponte di Wheatstone vicino all'idealità, per quanto riguarda l'accuratezza.

Purtroppo, l'incertezza intrinseca delle resistenze non è l'unica fonte di incertezza nelle misure mediante ponte di Wheatstone: un effetto è causato dal montaggio del ponte, ossia dalle resistenze di contatto tra i vari rami; un altro, è causato dall'effetto Seebeck: un elemento, riscaldato, genera una tensione spesso non trascurabile.

Per quanto riguarda l'effetto Seebeck, se nel ponte si ha a che fare con contatti di materiali di tipo diverso (per esempio resistori di materiale diverso dai fili), a temperature diverse tra loro, si generano delle tensioni dette forze termoelettromotrici, esprimibili mediante una funzione $f(\theta_1 - \theta_2)$, dove θ_1 e θ_2 sono le temperature dei due materiali a contatto. Essendo abbastanza complicata questa funzione f, si sceglie spesso di svilupparla, linearizzandola o considerando contributi superiori, ottenendo:

$$e_{FTEM} = f(\theta_1 - \theta_2) \approx \alpha(\theta_1 - \theta_2) + \beta(\theta_1 - \theta_2) + ...$$

Dove $\alpha \sim 5 \div 50\frac{\mu V}{K}$, e spesso β è trascurabile.
Esistono due approcci per eliminare questo apporto di tensione parassita:

- Approccio pratico: cercare di portare ogni componente del circuito alla stessa temperatura θ, in modo da evitare il generarsi di forze termoelettromotrici pur avendo materiali diversi tra loro.
- Approccio teorico: mediante uno studio della potenza dissipata da ogni singola resistenza, sommando in un unico contributo e_{TEM} tutta la forza termoelettromotrice, la si può mettere in serie al generatore E in ingresso, e così quantificare l'errore di consumo, ottenendo la possibilità di eliminarlo.

Per quanto riguarda le resistenze di contatto, supponendole per semplificazione tutte uguali a R_p, possiamo pensare che, con una misura 2-wired, si misuri un oggetto del tipo:

$$R_x = \frac{(R_a + 2R_p)(R_c + 2R_p)}{R_b + 2R_p}$$

Introducendo dunque un effetto sgradevole alla nostra misura, ed una notevole incertezza, se R_p non è trascurabile rispetto ad ognuna delle resistenze. Se così non fosse, è il caso di cercare rimedi alternativi, quali configurazioni differenti del ponte, in grado di favorire una resistenza del tipo 4-wired, in grado di attenuare le resistenze di contatto. Esistono alcune topologie, utilizzando resistori a 4 morsetti, in grado di ridurre anche notevolmente queste resistenze parassite.

6.2 Ponte di Wheatstone fuori equilibrio

Abbiamo sinora parlato del ponte di Wheatstone considerandolo sempre in equilibrio, ossia utilizzandolo solo nell'intorno di uno zero di trasmissione, con la tensione $V_u \simeq 0$, linearizzandone la caratteristica di uscita in questo punto di lavoro.

In realtà il ponte di Wheatstone è alla base di molte idee nate in seguito per misure di diverso tipo, quali di impedenze (ponte di Shering), o altro. Un'idea che intendiamo osservare è quella di sfruttare il ponte come trasduttore resistenza-tensione: anzichè considerare la caratteristica del ponte linearizzata in un intorno dello zero, si decide di utilizzarne una porzione più elevata, studiando la caratteristica del ponte al variare della resistenza, e quindi collegando il ponte a sensori di tipo resistivo. Utilizzando dunque resistenze variabili con la temperatura, con la deformazione, con lo spostamento, è possibile tradurre una variazione di resistenza in una variazione di tensione di uscita al ponte, V_u.

Il problema del ponte è il fatto che la caratteristica, solitamente, è molto dolce, e dunque risulta difficile definire precisamente la sensibilità del trasduttore, basato sulla linearizzazione della caratteristica del ponte in punti diversi da quello di equilibrio: il nostro scopo è quello di ottenere curve più ripide, con coefficienti angolari più marcati, in modo da essere meglio sensibili a strumenti di misura. Possiamo usare a questo punto diverse strategie, al fine di rendere più ripida la caratteristica, evidenziando meglio le variazioni delle uscite in particolari casistiche.

- Una strategia può essere quella di modificare il ponte, introducendo un amplificatore operazionale sull'uscita, amplificandola e quindi rendendo più ripida la nostra caratteristica. Supponiamo che l'operazionale abbia guadagno infinito, e dunque entri una corrente nulla ai suoi morsetti, che si trovano allo stesso potenziale. Considerando ciò, facendo alcuni conti di elettrotecnica, si verifica che:

  $$V_u = K_0 + K_1 R$$

  Dove K_0 e K_1 sono legati all'introduzione dell'operazionale, ed R è il resistore variabile, il sensore resistivo che introduciamo nel ponte per trasdurre la grandezza che percepisce in tensione, amplificata dall'operazionale.

- Strategia alternativa può essere quella di utilizzare più sensori della stessa grandezza, su lati opposti del ponte: in questo modo, aumenta (nel caso di resistori tutti uguali, raddoppia addirittura) la sensibilità del ponte, e così si riesce ad ottenere comunque un risultato soddisfacente.

Un esempio pratico di sensore collegato ad un ponte di Wheatstone è l'estensimetro (strain gauge): si tratta di un dispositivo in grado di variare la propria resistenza, al variare della propria geometria: all'interno del dispositivo vi è una pista resistiva che varia in lunghezza e larghezza assieme all'intera piastra e così, seguendo le leggi di Ohm, varia la propria resistenza. Collegato al ponte di Wheatstone, la variazione della resistenza di questo dispositivo è in grado di provocare variazioni di tensione in uscita al ponte, e quindi di misurare in tensione la deformazione della piastra.

Capitolo 7

Misure di Impedenze

Ben note sono le tecniche di misura di resistori, mediante ponti, tecniche amperometriche, o tecniche di altra categoria; meno note e sicuramente più delicate sono le tecniche di misura di impedenze, ossia di elementi con parte resistiva (reale), e reattiva (immaginaria): partendo dalle conoscenze sulla sola parte reale, e sulla Teoria dei Segnali, sono nate tecniche in grado di studiare e misurare le impedenze, inventando strumenti basati su concetti fisici molto particolari. Qua saranno esposte sostanzialmente tre tecniche, basate su tre modi di porsi dinnanzi all'argomento diversi tra loro:

- Ponte di Schering: un ponte di misura simile concettualmente al ponte di Wheatstone, ma alimentato in corrente alternata, e con un certo numero di accorgimenti;
- Q-metro: dispositivo in grado di misurare impedenze mediante rilevamento della condizione di risonanza del circuito;
- Impedenzimetro vettoriale: basato sul metodo voltamperometrico di misura

7.1 Ponti di Impedenze

La progettazione del ponte di Wheatstone ha dato luce ad un insieme di sistemi di misura che vanno anche molto al di là della semplice misura di resistenza: mediante resistori variabili con determinati parametri, si è arrivati a misurare, mediante il ponte di Wheatstone, temperature, contrazioni o dilatazioni fisiche, e molti altri parametri. Tutto ciò si basa, in modo indiretto, ad una banale misura di resistenza, mediante un ponte alimentato in corrente continua (DC).

A partire da questo ponte, alcune sue evoluzioni sono in grado di misurare non solo resistenze, ma anche eventuali componenti reattive, mediante l'alimentazione del nuovo ponte con un generatore di corrente alternata (AC), sfruttando un metodo di zero (come nell'uso più classico del ponte di Wheatstone), ossia facendo annullare la tensione di uscita del ponte.

Nel ponte di Wheatstone, alimentato in DC, era sufficiente, al fine di ottenere la condizione di ponte in equilibrio (e quindi tensione in uscita nulla), modificare il valore di una resistenza, fino a rilevare il valore in grado di annullare la tensione in uscita; utilizzando una tensione/corrente variabile nel tempo, dunque una AC, la variazione di un elemento resistivo non è sufficiente per introdurre una condizione di equilibrio, poichè non è possibile compensare eventuali componenti reattive presenti all'interno del circuito (contenute nel campione, nell'incognita, o anche negli altri resistori, piuttosto che nei cavi): non basta più dunque poter variare esclusivamente componenti resistive, ossia componenti reali, se vogliamo pensarle in un contesto matematico: abbiamo bisogno di variare anche la componente reattiva, ossia la componente immaginaria, introdotta da capacità o induttanze contenute nel sistema di misura.

Consideriamo, per ipotesi, di avere dunque un comportamento più ideale possibile dei componenti reattivi, soprattutto parlando di condensatori: costruendo condensatori in aria, ossia che utilizzano l'aria come dielettrico (la cui costante dielettrica relativa a quella del vuoto è molto vicina ad 1), si ha una minore dispersione di potenza, riducendo errori di consumo e quindi perdite di vario tipo; qualcosa di simile si realizza anche per gli induttori, sempre costruiti in aria.

Come elemento variabile non utilizzeremo più dunque un resistore puro, bensì un condensatore a capacità variabile, tarabile in termini di variazioni: parlare di taratura di capacità costanti non è assolutamente facile, poichè, per quanto riguarda correnti alternate ed elementi reattivi, è molto facile aver a che fare con capacità parassite C_p; poichè per utilizzare capacità variabili utilizziamo sempre lo stesso condensatore, del quale modifichiamo la geometria mediante artifici di vario genere, possiamo considerare correlati i due valori di capacità: dati due valori di capacità assunti dal nostro condensatore variabile, C_1 e C_2, in ciascuno di essi vi sarà un termine parassita; la cosa veramente interessante è che, poichè possiamo considerare correlati C_1 e C_2, allora possiamo considerare:

$$C_1 = C_{v,1} + C_p$$

$$C_2 = C_{v,2} + C_p$$

Dove C_p, ossia il termine parassita, è uguale in entrambe le capacità!

In questo modo, pur avendo una notevole incertezza sul singolo valore di capacità, sulla differenza avremo un'incertezza molto minore: considerando infatti la differenza delle capacità, ossia la variazione dalla capacità C_1 alla capacità C_2, si ottiene:

$$\Delta C = C_2 - C_1 = C_{v,1} + C_p - (C_{v,2} + C_p) = C_{v,1} - C_{v,2}$$

In questo modo avremo eliminato una notevole incertezza, dovuta al termine parassita C_p, considerando dunque non valori stabili di capacità, quanto variazioni tra due valori.

Esistono diversi tipi di ponte basati sui principi finora accennati utilizzati per la misura di impedenze a partire dall'annullamento di una tensione misurata: il ponte di Schering, il cui nome deriva dal suo inventore (il fisico tedesco Harald Schering) è solo una di queste topologie utilizzate.

In questo caso, mediante la variazione di capacità tarate si potrà ottenere l'annullamento della tensione di uscita del ponte, ottenendo, mediante alcuni accorgimenti che discuteremo in seguito, buone prestazioni fino a frequenze di alcune decine di MHz.

7.1.1 Ponte di Schering

Vediamo che il ponte di Schering ha una topologia come quella qui presentata:

La condizione di equilibrio ora non dipende solo più puramente dalla resistenza di uscita, ma dalla frequenza dell'alimentazione: al variare di questa, varieranno le impedenze, modificate dagli elementi reattivi del circuito (quali le capacità, in questo caso). Mediante la variazione delle capacità variabili (come già detto realizzate mediante condensatori variabili con la loro geometria, modificabili mediante una manopola esterna), nel nostro caso C_1 e C_3, sarà possibile portare il ponte ad una situazione di equilibrio.

La condizione di equilibrio dunque in questo caso sarà:

$$Z_1 Z_3 = Z_2 Z_4$$

Questo ponte non contiene ancora l'impedenza incognita che intendiamo misurare: esistono diverse scelte, a seconda del tipo di impedenza che ci troviamo a misurare: è possibile introdurre, nella topologia che stiamo utilizzando, l'impedenza Z_x o in parallelo alla capacità variabile C_1, o in serie alla capacità variabile C_3. La scelta di uno di questi due metodi dipende da quanto l'introduzione dell'impedenza in uno piuttosto che in un altro punto sbilancia il ponte: si desidera di solito uno sbilanciamento non troppo elevato, in modo da poter essere ribilanciato mediante la sola variazione delle due capacità C_1 e C_3: se lo sbilanciamento è invece troppo elevato rispetto al range di valori che possono assumere queste due capacità, significa che è necessario utilizzare l'altra strada (e se neanche questa fosse idonea, sostituire le capacità variabili con capacità in grado di assumere un range più idoneo di valori rispetto alla misura che stiamo effettuando).

Vediamo come dobbiamo dunque comportarci, rispetto a ciascuno dei due casi.

Inserzione di Z_x in serie a C_3

Dato il seguente schema di lavoro:

La misura dell'impedenza $Z_x = R_x + jX_x$ si esegue seguendo due fondamentali step:

1. Si equilibra il ponte prima dell'introduzione della Z_x, regolando le capacità variabili C_1 e C_3, fino a ottenere $V_u = 0$, e leggendo i valori di $C_{1,1}$ e $C_{3,1}$;

2. Si inserisce Z_x nel ponte, e si riazzera variando nuovamente le capacità variabili, leggendo i valori $C_{1,2}$ e $C_{3,2}$.

Imponendo le condizioni di equilibrio, possiamo ricavare, mediante semplici conti di Elettrotecnica, che la resistenza e la reattanza valgono:

$$R_x = \frac{R_4}{C_2}(C_{1,2} - C_{1,1})$$

$$jX_x = \frac{1}{j\omega}\left(\frac{1}{C_{3,1}} - \frac{1}{C_{3,1}}\right)$$

Presentiamo una piccola nota: quando è necessario, allora si senta assolutamente libertà di lavorare con il sistema appena esposto; per ragioni di montaggio, però, si vuole ricordare che senza dubbio considerare un montaggio in serie è più complicato che considerarne uno in parallelo: montando un circuito in serie è necessario spezzare un collegamento, ed aggiungerne uno nuovo per mettere appunto in serie le impedenze; il parallelo è senza dubbio più semplice, poichè per realizzarlo è sufficiente collegare agli stessi morsetti dell'impedenza del circuito la nostra incognita.

Inserzione di Z_x in parallelo a C_1

Lo schema del circuito di lavoro in questo caso sarà il seguente:

Poichè lavoriamo con dei paralleli, sarà opportuno utilizzare, al posto dei parametri resistenza R e reattanza X, i parametri ammettenza G e suscettanza Y, in modo da avere conti più facili da fare, misurando dunque questa volta l'impedenza $Z_x = G_x + jY_x$. Il procedimento è abbastanza simile a quello precedentemente esposto nel caso della serie:

1. Equilibrare il ponte senza aver ancora introdotto l'impedenza incognita Z_x, regolando le capacità variabili C_1 e C_3, leggendo i valori $C_{1,1}$ e $C_{3,1}$;
2. Inserire in parallelo alla capacità C_1 l'impedenza incognita C_1, ed equilibrare il ponte, regolando come al solito le capacità variabili, ottenendo i valori $C_{1,2}$ e $C_{3,2}$.

Mediante calcoli di Elettrotecnica, si otterrà:

$$G_x = \frac{C_2}{R_4}\left(\frac{1}{C_{3,2}} - \frac{1}{C_{3,1}}\right)$$

$$jY_x = j\omega(C_{1,1} - C_{1,2})$$

7.1.2 Schermatura del Ponte di Schering

Abbiamo descritto come procedere con la misurazione delle impedenze in due casi specifici, non tenendo per ora conto di un fatto molto importante: misurando le impedenze mediante le capacità finora descritte, si deve tener conto che nella nostra misura è presente un contributo causato dalle capacità parassite presenti nel ponte. Abbiamo detto che misurando variazioni possiamo considerare inalterato il contributo della C_p prima introdotta, ma ciò non è del tutto vero, poichè, durante le due fasi di misurazioni (sia nel primo che nel secondo caso) le capacità parassite variano.

Altro fatto da correggere è intrinseco alla topologia del circuito: sia il generatore di segnale (sul ponte) G che il rilevatore R dovrebbero avere un morsetto a 0 V, in modo da effettuare l'emissione di segnale e la misurazione a partire dallo stesso sistema di riferimento, e quindi avere un sistema ben definito ed una misurazione ben effettuata, studiando grandezze con lo stesso 0 V.

Il metodo di misura constaterà sempre di due fasi, come prima descritte, ma si cercherà di eliminare le capacità parassite dal calcolo finale, rendendole costanti mediante tecniche di schermatura, che non permetteran loro di variare da una fase all'altra del processo di misurazione.

Come si effettua il procedimento di schermatura dunque? Molto semplice: ogni singola impedenza viene introdotta in una scatola metallica, in modo che il ponte non abbia problemi causati da fenomeni esterni; se le dimensioni di questa scatola sono sufficientemente piccole rispetto alla lunghezza d'onda del segnale di alimentazione, di eccitazione, possiamo pensare che la scatola sia equipotenziale. Ciascuno di questi contenitori è collegato allo 0 V, spostando di fatto le impedenze variabili, in una zona che non riguarda la nostra misura: poichè abbiamo uno schermo equipotenziale, e poichè la corrente si calcola come derivata della carica nel tempo, la carica è legata alla tensione ma la tensione è nulla, quindi non si ha di fatto corrente sullo schermo (migliorando quindi anche la sicurezza dell'operatore di misura che, pur introducendo un'impedenza verso terra, non modifica l'equilibrio del sistema).

Lo schema del ponte di Schering schermato sarà dunque il seguente:

In condizioni di equilibrio, la tensione ai capi del rilevatore R sarà 0 V, e dunque avremo tutti gli schermi equipotenziali e ad un potenziale pari a 0 V, avendo così stabilizzato le capacità parassite, poichè non sono soggette a tensione.

Rimane un problema: se il ponte è all'equilibrio, il rilevatore ha effettivamente un morsetto a 0 V; i nodi del generatore G tuttavia sono ancora flottanti, ossia non sono collegati a 0 V, e ciò potrebbe introdurre impedenze parassite di contatto, $Z_C e Z_D$: queste andrebbero a mettersi a 0 V in parallelo alle capacità presenti sui due rami.

Ciò richiede un'ulteriore modifica del nostro circuito, modificando l'unico elemento finora lasciato inalterato: il generatore.

Per schermare il generatore, possiamo spostare il problema, introducendo un trasformatore: sulla diagonale del ponte, CD, introduciamo, al posto che il generatore, un induttore accoppiato ad un altro, collegato al generatore, con uno dei morsetti a 0 V. L'effetto trasformatore permetterà al generatore di alimentare comunque la diagonale CD del ponte, non variando di fatto in modo considerevole la configurazione del ponte di Schering. Abbiamo però in realtà solo traslato il problema, ossia modificato le capacità parassite Z_C e Z_D, che non sono però ancora definite, poichè sono comunque ai capi del trasformatore, rendendone imprevedibili le prestazioni. Eliminarle però sarà facile, schermando il trasformatore: le induzioni causate dalla rete di alimentazione non potranno più provocare problemi, ed avremo quindi, in altre parole, ben definito le ultime capacità parassite presenti nella rete.

Come si realizza la schermatura del trasformatore? Semplice: inserendo uno schermo elettrostatico, ossia una parete metallica, non chiusa su se stessa, ma collegata uno ai capi del primario, uno ai capi del secondario: da un lato lo schermo elettrostatico deve bloccare l'influenza di campi elettrici esterni in modo da evitare interferenze esterne, ma dall'altro lato non deve circolare corrente all'interno, in modo da evitar di alimentare le impedenze e così funzionando esclusivamente da schermo, da riferimento di potenziale, non da conduttore.

7.2 Q-metro

Con il ponte di Schering la misura si basava su di una tecnica di zero: la condizione di equilibrio che permetteva di rilevare parametri mediante i quali si potevano calcolare le parti reali ed immaginarie dell'impedenza, era una condizione di zero di trasmissione, ossia un'uscita nulla; alla base del qmetro è la ricerca di un'altra particolare condizione del circuito, ossia la condizione di risonanza.

7.2.1 Chiarimenti sul concetto di risonanza

Fisicamente, per risonanza elettrica, ossia risonanza considerata nel contesto di un circuito elettrico/elettronico, si intende la condizione che si viene a verificare quando, in una certa frequenza ω_0 di un segnale eccitante il circuito, l'energia del circuito si trova ad oscillare tra il campo magnetico di un induttore, ed il campo elettrico di una capacità. Modellando il circuito elettrico con una capacità e un'induttanza, possiamo pensare a ciò: la variazione del campo magnetico presente nell'induttore genera una corrente che va a caricare il condensatore; a sua volta, questo, scaricandosi, genererà la corrente che indurrà per autoinduzione nell'induttanza il generarsi di un campo magnetico, continuando, armonicamente, con questo ciclo di oscillazioni.

Considerando il nostro circuito, abbiamo a disposizione due possibilità: capacità e induttanza in serie, oppure in parallelo; in questi due casi, il comportamento del circuito, sarà esattamente duale, poichè l'impedenza equivalente delle due reattanze, in condizione di risonanza, sarà minima in serie, e massima in parallelo. In questo secondo caso, che noi tratteremo in modo più approfondito per quanto riguarda i nostri sistemi di misura, significherà avere il massimo della funzione di trasferimento, e quindi il massimo di $|H(j\omega)|$. Mediante calcoli, è possibile ricavare le espressioni operative delle impedenze equivalenti in caso di serie (Z_s) o parallelo (Z_p):

$$Z_s = R + j\left(\omega L - \frac{1}{\omega C}\right)$$

$$Z_p = \frac{j\omega L}{1 - \omega^2 LC}$$

Dove ω è la pulsazione, L l'induttanza, C la capacità, e R la resistenza parassita, la resistenza di consumo del circuito (che discuteremo meglio parlando del metodo di misura).

Quando si ha dunque il contributo della reattanza induttiva uguale a quello della reattanza capacitiva, si ha la condizione di risonanza, legata ad una determinata frequenza detta frequenza di risonanza, ω_0. Sia nel caso di risuonatori serie che di risuonatori parallelo, la frequenza di risonanza vale:

$$\omega_0 = \frac{1}{\sqrt{LC}}$$

Finora abbiamo considerato tutto ideale, ossia senza considerare le perdite dei componenti utilizzati per realizzare il circuito risuonatore; considerando R_P la resistenza contenente tutti gli errori di consumo (che discuteremo in seguito), e G la sua ammettenza, legata ad essa dalla formula:

$$G = \frac{1}{R_P}$$

Si definisce il fattore di qualità, o fattore di merito, Q, come rapporto tra la potenza reattiva e la potenza attiva del circuito risonante in una determinata uscita; svolgendo direttamente i conti, separando Q_s nel caso serie, e Q_p nel caso parallelo:

$$Q_s = \frac{\omega_0 L}{R_P}$$

$$Q_p = \frac{R}{\omega_0 L}$$

Consideriamo nel nostro caso solo circuiti risuonatori parallelo, e quindi $Q \triangleq Q_p$; a seconda delle necessità, usando l'ammettenza equivalente di consumo G, è possibile ricavare le espressioni equivalenti di Q dipendenti sia da capacità che da induttanza:

$$Q = \frac{1}{\omega_0 LG} = \frac{\omega_0 C}{G} = \omega_0 C R_P = \frac{R_P}{\omega_0 L}$$

7.2.2 Misura del fattore di qualità Q

Come si può intuire dal nome dello strumento, il qmetro è uno strumento che, mediante la misura del fattore di qualità, Q, è in grado di determinare il valore di un'impedenza. Il nostro scopo a questo punto è determinare un certo numero di metodi di misura di questo Q, al fine di poter misurare indirettamente l'impedenza; a seconda delle situazioni in cui ci troveremo, potremo adottare un metodo piuttosto di un altro, e così poter eliminare problemi diversi in modo appropriato per ciascuno di essi.

Considerando ad esempio in ambito filtristico il circuito risonante, esso si può pensare come un filtro passa-banda, dove la banda passante è centrata nella frequenza di risonanza, $f_0 = 2\pi\omega_0$;

Si può quindi relazionare il fattore di qualità Q del circuito con la larghezza di banda passante B, calcolata in un qualche modo (banda equivalente, banda a -3 dB, banda null-to-null o altre; di solito si considera la banda a -3 dB):

$$Q = \frac{\omega_0}{2\pi B}$$

Come si deve procedere in pratica: si cerca di determinare il valore di ω_0, se possibile mediante la formula prima enunciata (se si riesce a scomporre tutte le capacità e le induttanze in solo un induttore ed un condensatore), oppure cercando il massimo della funzione di trasferimento, modificando manualmente la frequenza del generatore di tensione $E(\omega_0)$. Mediante altre modifiche della frequenza, si cerca il valore della banda a -3 dB, e così si riesce a determinare il valore del Q, misurando il massimo di $V(\omega)$ nel dominio della frequenza, e così ω_0 e banda.

Si può pensare anche ad un metodo alternativo di misura del Q, basato sullo studio della funzione di risposta all'impulso, e quindi su di uno studio nel dominio del tempo, del nostro circuito di misura: quando non siamo in condizioni di risonanza, infatti, l'andamento nel tempo della tensione ai capi di uno degli elementi reattivi, o L o C, è di fatto una cisoide, ossia una sinusoide dotata di smorzamento esponenziale. In una cisoide, abbiamo dunque un decadimento esponenziale delle ampiezze (possiamo dire che le cisoidi siano sinusoidi che seguono come inviluppo un esponenziale reale), ma la distanza sull'asse dei tempi di due massimi (o minimi) relativi di fatto non varia.

Si può contare il periodo di cicli, N, per la durata del transitorio, identificata dalla costante di tempo τ. Dato t_0 ogni distanza tra due massimi, e dato τ il tempo che ci impiega, dall'inizio della prima oscillazione, l'ampiezza a normalizzarsi per e, dove e è il numero di Nepero, si dimostra che:

$$\frac{\tau}{t_0} = \frac{Q}{\pi} \simeq N$$

Abbiamo in questo modo trovato una relazione utile ai fini della nostra misura di Q: possiamo infatti dire, ribaltando l'ultima equazione:

$$Q = N\pi$$

Questi sono i principi alla base del qmetro.

7.2.3 Realizzazione di un Q-metro

Il circuito alla base del nostro strumento di misura deve essere un circuito risonante a basse perdite (nella fatispecie nella maggior parte dei casi esse derivano dall'induttanza L). Poichè è sempre valido il principio di equivalenza, è possibile trasformare un normale circuito RLC serie, composto dal resistore R_s, dalla capacità C_s, e dall'induttore L_s, in un equivalente più facile da studiare, con una particolarità: considerare tutti gli elementi ideali, concentrando le perdite causate dagli elementi reattivi non ideali in una singola ammettenza G, in parallelo ad una nuova capacità C_p.

Come sappiamo dalla teoria dei circuiti risonanti, quando il circuito si trova alla pulsazione di risonanza ω_0, valgono le seguenti relazioni:

$$\omega_0 L_p = \frac{1}{\omega_0 C_p}$$

$$Q = \frac{1}{\omega_0 L_p G} = \frac{\omega_0 C_p}{G}$$

$$V_0 = EQ$$

Dove per V_0 intendiamo la tensione massima, la tensione ai capi di una delle reattanze, quando ci troviamo in condizioni di risonanza, ossia alla pulsazione ω_0.

Se non avessimo effetti di perdita, avremmo una situazione di questo genere:

L'ammettenza G serve semplicemente a modellizzare tutti i fenomeni di perdita del circuito; per circuito intendiamo ovviamente sia i componenti utilizzati per la misura, sia i misurandi: qualsiasi elemento appartenente al circuito introducente una qualche resistenza parassita va considerato elemento di non idealità, e quindi aggregato in G. Se avessimo un qmetro ideale, e quindi non avessimo nessun effetto di perdita in esso, potremmo pensare che $G \to \infty$, e quindi potremmo pensare che, per $\omega \to \omega_0$, si avrebbe un asintoto verticale, e quindi $V(\omega) \to \infty$. Naturalmente ciò è assolutamente impossibile, in quanto non si può avere un Q infinito.

Nella realtà, è possibile realizzare induttori e condensatori di buona qualità, utilizzando il migliore dei dielettrici: l'aria. Le perdite sarebbero ridotte dagli effetti isolanti dell'aria, e quindi si potrebbero ottenere in questa maniera induttori e condensatori relativamente piccoli (induttori da circa 10 nH fino a qualche centinaio di μH, e condensatori variabili da qualche decina a qualche centinaio di pF).

Studiando dunque componenti con capacità e induttanze di questi ordini di grandezza, possiamo capire che la frequenza di risonanza si aggirerà da qualche centinaio di kHz a qualche decina di MHz, e quindi non si potran studiare componenti comportanti una frequenza di risonanza al di fuori di quel range.

La realtà non ci viene incontro: per quanto buoni siano gli induttori ed i condensatori realizzati, per quanto elevata sia l'impedenza in ingresso del voltmetro utilizzato come rilevatore del fattore di qualità Q, e per quanto bassa sia l'impedenza del generatore di tensione polarizzante il nostro circuito, sicuramente non avremo, come nel caso ideale, un asintoto verticale della funzione di trasferimento in prossimità della pulsazione di risonanza ω_0, bensì solo un massimo assoluto, finito.

La G in parallelo a tutti gli altri elementi è non nulla, poichè racchiude tutti gli effetti di non idealità prima esposti; a queste condizioni, possiamo considerare ideali tutti gli altri componenti, e quindi pensare che, nel circuito equivalente da noi introdotto, le non-idealità siano presenti solo e soltanto nell'ammettenza in parallelo, G.

In una misura di questo tipo, come già suggerito precedentemente, abbiamo fondamentalmente due gradi di libertà, due elementi variabili sui quali lavorare: la frequenza di pulsazione del generatore E, ω, che da luogo quindi a $E(\omega)$, e C_V: quest'ultimo è l'elemento reattivo che siamo in grado di variare, ai fini di ottenere la frequenza di risonanza imposta dal resto del circuito, e dalla polarizzazione.

Solitamente, si varia solo uno dei due parametri: nella maggior parte dei casi, infatti, capita di studiare un'impedenza reagente ad una certa frequenza imposta dal generatore; l'unico grado di libert sar dunque la capacità C_v, mediante la cui variazione potremo portare il circuito alla risonanza, modificando la sola reattanza capacitiva. Altro problema che può sussistere, ma molto raramente studiabile, è considerare sia l'induttanza L_p che la capacità C_v costanti, mentre la frequenza della polarizzazione $E(\omega)$ variabile; a questo punto è possibile variarla fino a trovare il massimo della funzione di trasferimento $H(j\omega)$, in prossimità della frequenza di risonanza del circuito, ω_0.

7.2.4 Misura di una impedenza mediante Q-metro

Consideriamo, al fine di poter arrivare più facilmente alla determinazione della misura di impedenze qualsiasi, tre casi semplici, che poi potremo unire insieme determinando il caso più generale: un'impedenza puramente capacitiva, C_x, un'impedenza puramente induttiva, L_x, e un'impedenza puramente resistiva, R_x (o, in alternativa, una conduttanza G_x uguale al reciproco di R_x)

Alla base delle misure che effettueremo, sarà la determinazione della condizione di risonanza, al fine di poter considerar valide le ben note relazioni:

$$\omega_0 L_p = \frac{1}{\omega_0 C_p}$$

$$Q = \frac{1}{\omega_0 L_p G} = \frac{\omega_0 C_p}{G}$$

Dove L_p e C_p sono i valori delle reattanze equivalenti rispettivamente induttiva e capacitiva del circuito utilizzato come Q-metro. La condizione di risonanza è rilevata mediante un rilevatore M, un voltmetro, con però la scala graduata tarata sul Q e non sulle tensioni V: poichè consideriamo nei nostri esempi un circuito RLC parallelo, la posizione del Q massimo sarà indicatrice di una condizione di risonanza.

Consideriamo ora, a partire dalle ipotesi finora proposte, i tre casi appena citati.

Misura di una capacità pura C_x

Supponiamo di voler misurare un'impedenza puramente reattiva, capacitiva; sarà necessario seguire alcune fasi, al fine di ultimare il processo di misura:

1. Realizzare il circuito del Q-metro, senza introdurre l'elemento incognito, scegliendo un induttore adatto alla misura che si intende effettuare: la dinamica del condensatore variabile non è infatti infinita, ma è in grado di compensare lo sbilanciamento introdotto dal misurando solo entro un certo range di valori; per questo motivo, l'induttore L_p dovrà essere tale da permettere, una volta introdotto l'elemento incognito, di tornare alla frequenza di risonanza mediante la sola variazione del condensatore variabile C_V. La frequenza di risonanza si definisce mediante la regolazione della frequenza del generatore alimentante il circuito, e sarà chiamata ω_0; realizzato il circuito e introdotta l'alimentazione a pulsazione ω_0, mediante la variazione dei valori di C_V, troveremo il valore $C_{V,1}$, ossia il valore della capacità variabile tale per cui l'ampiezza della tensione misurata ai capi del misuratore M sia massima, e quindi il Q massimo (in altre parole, si porta in risonanza il circuito RLC); in queste condizioni, valgono le relazioni:

$$\omega_0 L_p = \frac{1}{\omega_0 C_{V,1}}$$

$$Q = \frac{1}{\omega_0 L_p G} = \frac{\omega_0 C_{V,1}}{G}$$

2. Si introduce a questo punto la capacità incognita C_x, in parallelo al condensatore variabile C_V (attualmente con il valore $C_{V,1}$, ossia il valore che portava in risonanza il circuito prima dell'introduzione di C_x); il circuito andrà sicuramente fuori risonanza, poichè l'introduzione del misurando ha fatto variare il Q misurato, facendolo abbassare rispetto al valore precedente (che era il valore massimo, in quelle condizioni). Variando C_V fino a trovare il valore in grado di riportare nuovamente il circuito in condizione di risonanza, $C_{V,2}$, potremo ritenere valide le seguenti formule:

$$\omega_0 L_p = \frac{1}{\omega_0 C_{V,2} + C_x}$$

$$Q = \frac{1}{\omega_0 L_p G} = \frac{\omega_0 C_{V,2} + C_x}{G}$$

3. Confrontiamo le espressioni prima e dopo l'inserimento della capacità incognita: la pulsazione ω_0 di fatto non varia, in quanto non abbiamo in alcun modo modificato il generatore; inoltre, l'induttanza L_p, scelta all'inizio del primo punto, non sarà variata, poichè nessun parametro introdotto può modificarla. Abbiamo dunque due punti fermi sui quali basarci, e sul quale scrivere le nostre relazioni:

$$Q = \frac{1}{\omega_0 L_p G}$$

$$\omega_0 L_p = \frac{1}{\omega_0 C_{V,2} + C_x} = \frac{1}{\omega_0 C_{V,1}}$$

Da quest'ultima espressione, si vede, date le precedenti ipotesi, che:

$$C_{V,1} = C_{V,2} + C_x \Longrightarrow C_x = C_{V,1} - C_{V,2}$$

Dal momento che consideriamo tutto ideale, ossia pure reattanze, non saran aumentate le perdite, non sarà variata G, dunque possiamo dire che i due Q misurati dal rilevatore, non varino da prima a dopo l'inserimento del misurando. Si noti che, inoltre, se Q fosse basso, potrebbe dipendere anche da elementi puramente reattivi; consideriamo ora e sempre verificata l'ipotesi di Q piuttosto elevati, almeno superiori a 20 o 30.

Per quanto riguarda il calcolo delle incertezze, stimando con un modello di caso peggiore le incertezze sulle misure, possiamo dire che, poichè la relazione che permette di stabilire il valore di C_x è:

$$C_x = C_{V,1} - C_{V,2}$$

Per un modello di worst case, l'incertezza risultante sarà la somma delle incertezze assolute:

$$\delta C_x = \delta C_{V,1} - \delta C_{V,2}$$

Le incertezze appena presentate, $\delta C_{V,1}$ e $\delta C_{V,2}$, sono le incertezze con cui si conoscono i valori $C_{V,1}$ e $C_{V,2}$; esse possono essere piuttosto elevate, e dunque la somma delle due risultare non indifferente: per questo motivo, un modello di worst case, potrebbe essere inadatto, al fine di ottenere una stima realistica dell'incertezza, in questo ambito.

Come possiamo migliorare la nostra stima dell'incertezza? Studiamo il componente introducente indeterminazione, ossia il condensatore variabile C_V:

Si tratta di fatto di un condensatore a facce piane parallele, in cui la sovrapposizione delle facce viene regolata mediante una manopola (che possiamo modificare all'esterno); più i condensatori sono sovrapposti, ossia maggiore è l'area rappresentante l'intersezione tra le due lamine, e maggiore sarà la capacità risultante del condensatore variabile: di fatto quindi si modifica, mediante un artificio geometrico, la capacità mediante una manopola esterna. A seconda della geometria delle lamine (cilindriche piuttosto che circolari piuttosto che altre eventuali), è possibile variare le superfici linearmente piuttosto che logaritmicamente rispetto alla variazione della posizione della manopola indicatrice. Quello che si può intuire è che, con uno strumento di questo genere, le due capacità $C_{V,1}$ e $C_{V,2}$ abbiano una correlazione non nulla, e quindi, scomponendo l'incertezza assoluta δC_V in diversi contributi, si potrebbero notare contributi identici nei due valori. Per questo motivo, si è soliti tarare le capacità variabili di questo tipo semplicemente in termini di variazioni: utilizzando una taratura di questo genere, l'incertezza sulla variazione, ossia sulla differenza di valori, mediante un modello più realistico rispetto a quello di caso peggiore, risulta essere molto inferiore: il fatto che vi sia correlazione tra le due situazioni permette di eliminare il termine della correlazione, in modo da mantenere e sommare il solo termine di effettiva incertezza.

Possiamo considerare $C_{V,1}$ (e poi $C_{V,2}$) come scomposto in due termini: $C'_{V,1}$ e C_P , dove il secondo si può pensare come un contributo parassita, che introduce indeterminazione sul secondo termine; in questa maniera, formalizzando:

$$C_{V,1} = C'_{V,1} - C_P$$

$$C_{V,2} = C'_{V,2} - C_P$$

$$C_{V,1} - C_{V,2} = C'_{V,1} - C_P - (C'_{V,2} - C_P) = C'_{V,1} - C'_{V,2}$$

Sparendo i termini C_P, riusciamo ad eliminare buona parte dell'incertezza, cosa che non abbiam potuto fare nel modello di caso peggiore prima esposto.

Misura di un'induttanza pura L_x

Per misurare un'induttanza pura, possiamo scegliere due metodi, a seconda del tipo di induttanza incognita che dobbiamo studiare:

1. Metodo di sostituzione: al posto di utilizzare l'induttanza L_p, si sostituisce ad essa semplicemente la L_x, e si porta in risonanza il circuito: trovato il valore di capacità $C_{V,0}$ in grado di mandare in risonanza il circuito, possiamo utilizzare le solite relazioni, considerando però l'impedenza incognita L_x:

$$\omega_0 L_x = \frac{1}{\omega_0 C_{V,0}}$$

   Da qui:

$$L_x = \frac{1}{\omega_0^2 C_{V,0}}$$

2. Metodo simile a quello utilizzato con le capacità incognite C_x: prima di tutto si considera il circuito del Q-metro senza aver introdotto l'incognita L_x; a questo punto, si manda il circuito in risonanza, determinando il valore di capacità $C_{V,1}$ in grado di realizzare la condizione di risonanza. Potremo considerare dunque valida la relazione:

$$\omega_0 L_p = \frac{1}{\omega_0 C_{V,1}} \Longleftrightarrow \omega_0 C_{V,1} = \frac{1}{\omega_0 L_p}$$

Si introduce in parallelo all'induttore L_P l'incognita L_x (si noti che si sceglie di introdurre in parallelo esclusivamente per questioni di realizzazione circuitale: è molto più semplice realizzare un parallelo che una serie, nella pratica); mandiamo il circuito con L_x in risonanza, modificando il valore di capacità, fino ad arrivare alla condizione a noi utile, portandolo ad un valore $C_{V,2}$. A questo punto, si utilizza la seguente relazione:

$$\omega_0 L_{eq} = \frac{1}{\omega_0 C_{V,2}}$$

Dove:

$$L_{eq} = L_p \oplus L_x = \frac{L_p L_x}{L_p + L_x}$$

Sostituendo nella precedente:

$$\omega_0 C_{V,2} = \frac{1}{\omega_0 \frac{L_p L_x}{L_p + L_x}}$$

Sottraendo membro a membro ambo le equazioni dei due stati, prima e dopo l'introduzione di L_x, si può trovare che

$$L_x = \frac{1}{\omega_0^2 (C_{V,2} - C_{V,1})}$$

Abbiamo introdotto due metodi di misura di L_x: uno nuovo, ed uno fondamentalmente simile a quello già considerato nel caso di capacità: anche in questo il risultato sarà funzione di una differenza di valori di capacità di un condensatore variabile. La prima tecnica, si basava infatti sul determinare:

$$L_x = \frac{1}{\omega_0^2 C_{V,0}}$$

Ossia nel determinare un singolo valore di C_V, in grado di mandare in risonanza il circuito. La seconda tecnica si basa sul determinare:

$$L_x = \frac{1}{\omega_0^2 (C_{V,2} - C_{V,1})}$$

Ossia sul determinare la variazione $C_{V,2} - C_{V,1}$), e quindi una ΔC_V, ad una certa pulsazione dell'alimentazione ω_0.

Il secondo metodo è preferibile al primo in quanto il contributo delle capacità parassite di un condensatore variabile non è facilmente eliminabile, e dunque manco la taratura di quest'ultimo: misurando variazioni di capacità, è invece possibile eliminare i contributi parassiti, e quindi migliorare di molto la stima delle incertezze.

Misura di una resistenza pura R_x

La misura di una resistenza incognita, per quanto ideale, introduce una differenza rispetto ai due casi finora considerati: una reattanza pura può spostare lo stato del circuito dalla condizione di risonanza, ma non è in grado (a meno che Q sia molto basso) di modificare il valore di Q. Al contrario, l'introduzione di una resistenza è in grado di modificare il parametro G, ossia il parametro racchiudente l'insieme degli errori di consumo del circuito, e con esso quindi il fattore di qualità Q. Il procedimento di misura sarà simile a quello precedentemente visto per la capacità C_x, con alcune varianti, come ora meglio esporremo:

1. Una volta introdotto un L_p in grado di risuonare con uno dei valori assumibili da C_V, si manda il circuito in risonanza, rilevando il valore massimo del Q, che chiameremo Q_1, con un certo valore di capacità $C_{V,0}$. In queste condizioni, varranno le solite relazioni proprie della risonanza:

$$\omega_0 L_P = \frac{1}{\omega_0 C_{V,0}}$$

$$Q_1 = \frac{1}{\omega_0 L_p G} = \frac{\omega_0 C_p}{G}$$

2. Introduciamo nel circuito, in parallelo a G, la R_x, magari espressa come conduttanza G_x, ossia:

$$G_x = \frac{1}{R_x}$$

   A questo punto, il circuito si troverà ancora in stato di risonanza, poichè avremo modificato solo la parte resistiva del circuito, senza toccare la parte reattiva: questo perchè avremo sì un aumento delle perdite del circuito, ma non una variazione della reattanza equivalente. L'aumento delle perdite comporterà però una variazione del fattore di qualità Q del circuito, che diventerà, da Q_1, $Q_2 < Q_1$; poichè siamo in risonanza, valgono le seguenti espressioni:

$$\omega_0 L_p = \frac{1}{\omega_0 C_{V,0}}$$

$$Q_2 = \frac{1}{\omega_0 L_p (G + G_x)} = \frac{\omega_0 C_{V,0}}{G + G_x}$$

Ricaviamo G dalla prima delle due espressioni: essa infatti non varia, prima e dopo l'inserimento di G_x:

$$G = \frac{\omega_0 C_{V,0}}{Q_1}$$

Sostituendo nella seconda, mediante alcuni passaggi algebrici:

$$G_x = \frac{Q_1 - Q_2}{Q_1 Q_2} \omega_0 C_{V,0}$$

Leggendo quest'ultima formula, possiamo calcolare l'incertezza assoluta di caso peggiore, mediante le formule della stima delle incertezze. Esse saranno basate sul calcolare le incertezze sui valori di Q misurati. Il fattore di qualità Q si può determinare come massimo sulla scala del voltmetro utilizzato come rilevatore, ma anche in una maniera diversa: se disponessimo di tecniche avanzate, e molto precise, per la misura della frequenza di risonanza, e della banda a -3 dB rispetto ad essa, potremmo utilizzare un approccio filtristico, e quindi determinare Q mediante la relazione:

$$Q = \frac{f_0}{B}$$

Dove:

$$f_0 = \frac{\omega_0}{2\pi}$$

Senza sfruttare le tecniche di misura della frequenza di risonanza f_0, si devono calcolare le incertezze nel modo classico, usando le incertezze relative ed assolute; il problema è che non è facile determinare un rapporto di correlazione tra Q_1 e Q_2, quindi, poichè $\delta Q_1 \sim \delta Q_2 \sim 5\% \div 10\%$, l'incertezza equivalente valutata mediante worst case sarà molto grossa.

Misura di una generica Z_x

Possiamo a questo punto estendere i tre casi semplici appena presentati, unendoli, analizzando il caso più generale: l'introduzione di una generica impedenza Z_x. Sappiamo che, dato un circuito RLC in condizione di risonanza, e quindi in cui il fattore di guadagno Q è massimo, introducendo una pura reattanza si sposta il circuito dalla condizione di risonanza, senza comunque variare la curva e l'ampiezza del massimo fattore di guadagno raggiungibile, in condizione di risonanza; introducendo invece una resistenza pura, si abbassa il massimo fattore di guadagno, ma non si sposta il circuito dalla condizione di risonanza. Intuitivamente, introducendo una generica impedenza Z_x, avremo una combinazione dei due effetti dei due casi: avremo un abbassamento del massimo fattore di qualità del circuito risonante, e non ci troveremo più in condizione di risonanza. Una qualsiasi Z_x, è modellabile con un circuito equivalente parallelo o serie; supponiamo per comodità di aver a che fare con un circuito parallelo, come abbiamo sempre fatto finora (anche se questo discorso in questo caso è molto delicato da affrontare, come vedremo tra poco).

Si utilizza una tecnica analoga a quelle finora utilizzate, come vediamo ora:

1. Si porta il circuito RLC senza l'inserimento dell'impedenza in condizione di risonanza, misurando e segnando i valori fondamentali di capacità e fattore di qualità, $C_{V,1}$ e Q_1; saranno dunque come al solito valide le relazioni:

$$\omega_0 L_p = \frac{1}{\omega_0 C_{V,1}}$$

$$Q_1 = \frac{1}{\omega_0 L_p G} = \frac{\omega_0 C_{V,1}}{G}$$

2. Si inserisce (per esempio in parallelo) la Z_x al condensatore variabile, senza modificare la ω_0 del generatore alimentante il circuito; il circuito andrà quindi fuori dallo stato di risonanza, e contemporaneamente il suo valore massimo di fattore di qualità sarà ridotto. Variando C_V fino a trovare il valore $C_{V,2}$, capacità tale per cui il circuito si trova in condizioni di risonanza, si misura il nuovo massimo fattore di qualità, Q_2. A questo punto, date queste ipotesi, dobbiamo considerare casistiche diverse, a seconda del tipo di impedenza introdotta nel circuito; con

calcoli simili a quelli precedentemente utilizzati, sarà comunque possibile determinare parte reale (resistenza) ed immaginaria (reattanza) dell'impedenza che intendiamo misurare.

Trattiamo come primo caso un'impedenza capacitiva, introdotta in parallelo; le equazioni del circuito in queste condizioni avranno una forma del tipo:

$$\omega_0 L_p = \frac{1}{\omega_0(C_{V,2} + C_x)}$$

$$Q_2 = \frac{1}{\omega_0 L_p(G + G_x)}$$

A queste condizioni, riportandoci alle equazioni del punto 1 del nostro procedimento, possiamo ottenere che:

$$G_x = \frac{Q_1 - Q_2}{Q_1 Q_2}\omega_0 C_{V,1}$$

$$C_x = C_{V,1} - C_{V,2}$$

Si noti che il discorso della misura di impedenze, come già accennato, è molto più delicato rispetto a quello delle misure di pure resistenze e reattanze; la connessione più comune utilizzata è effettivamente quella in parallelo, per i motivi di realizzazione pratica precedentemente citati, ma non è sempre la più idonea: quando si inserisce un'impedenza contenente una parte resistiva molto piccola per esempio, messa in parallelo con G va a formare una G_{eq} molto superiore a G, aumentando quindi enormemente i consumi (come si può immaginare pensando ad un partitore di corrente). Se i consumi del circuito si alzano in maniera smisurata, il fattore di qualità Q rischia di abbassarsi troppo, provocando effetti molto gravi sullo studio che stiamo effettuando (quale influenzabilità, oltre che da parti resistive, da parti reattive, rendendo il Q-metro molto difficile da utilizzare come strumento di misura). Se la resistenza dunque fosse molto bassa, conviene introdurre Z_x in serie a G per esempio, al fine di non modificare di molto la situazione. Discorsi analoghi e già affrontati si dovrebbero fare per la parte reattiva: il condensatore variabile C_V può variare solo in un certo range di valori: non è detto che, introducendo in parallelo l'impedenza Z_x, si riesca a riportare con la sola variazione di C_V il circuito in una condizione di risonanza, e quindi anche in base a questo fatto è necessario scegliere la giusta topologia del circuito, ossia il giusto posizionamento dell'impedenza misuranda Z_x. Nella fatispecie, ci sono due casi, e due condizioni da rispettare tassativamente, al fine di non incappare in problemi come quello appena espresso; considerando dunque un'impedenza equivalente parallelo, dovranno essere rispettate le seguenti condizioni:

- Se l'impedenza è capacitiva, e abbiamo quindi una C_x, dovrà essere verificata la diseguaglianza:

$$C_x \leq C_{V,MAX} - C_{V,MIN}$$

- Se l'impedenza è induttiva, e abbiamo quindi una L_x, dovrà essere verificata la diseguaglianza:

$$L_x \geq \frac{1}{\omega_0^2(C_{V,MAX} - C_{V,MIN})}$$

Capitolo 8

Misure di potenza

Un parametro di cui spesso non si tiene completamente conto, effettuando misure di tensioni o correnti ai capi di un elemento appartenente ad un circuito, è la frequenza del segnale che si studia nel circuito (per esempio nel caso di amplificatori o attenuatori). A diverse frequenze il circuito risponde in maniera diversa ad un segnale, come vedremo tra poco; si suol classificare dunque le misure in due fondamentali categorie:

- Misure a BF (ossia a Bassa Frequenza): da 0 fino a un centinaio circa di kHz, possiamo considerare di trovarci ad una bassa frequenza;
- Misure a RF (ossia a Radio Frequenza): da alcune centinaia di kHz fino a parlare di onde millimetriche, ossia la cui lunghezza d'onda λ è dell'ordine dei millimetri.

Non parleremo di microonde, ossia di onde la cui λ è dell'ordine del μm; si sappia comunque che, variando la frequenza, il comportamento del circuito continua a variare, presentando effetti sempre più sgradevoli man mano che si aumenta con la frequenza.

Nell'Ingegneria Elettronica e delle Telecomunicazioni, dove si ha a che fare con segnali, ossia con entità dalla potenza molto ridotta, si studia di solito esclusivamente la potenza attiva del segnale, ossia la potenza collegabile concettualmente alla sola parte reale dell'intera potenza. Solitamente, il range di valori di potenza che si arriva a studiare parte dai nanowatt nW fino ad arrivare a qualche decina di watt, W. Solitamente, le misure di potenza saranno effettuate in maniera indiretta, ossia non misurando direttamente il parametro potenza, bensì tensioni e/o correnti. Questo, almeno, è vero in condizioni di bassa frequenza: ad alta frequenza, come nel caso delle radiofrequenze, si effettua la misura di potenza in base a misure di tipo termico, ossia si misura tutto in termini di potenza dissipata da elementi resistivi.

Si noti che spesso in Ingegneria Industriale, nella fatispecie Ingegneria Elettrica, i discorsi che abbiamo appena fatto e che faremo non valgono: si è soliti studiare anche la potenza reattiva, e diversi ordini di grandezza sopra quelli ai quali noi ci limiteremo.

Le misure di potenza sono meno accurate di altre grandezze precedentemente viste quali corrente o tensione, poichè, essendo misurate in maniera indiretta, coinvolgono diverse grandezze e quindi anche le grandezze di influenza sulle grandezze di misura, accrescendo notevolmente l'incertezza risultante sulla misura finale.

8.1 Misure di potenza in bassa frequenza

Le misure a bassa frequenza non sono molto complicate da effettuare: supponiamo di avere un circuito, del quale conosciamo, mediante specifiche, il valore della resistenza di carico R_L (con relativa incertezza); misurando semplicemente la tensione efficace ai suoi capi, V_{eff}, potremo trovare la potenza semplicemente mediante la relazione:

$$P_L = \frac{V_{eff}}{R_L}$$

Come mai parliamo di grandezze efficaci? La risposta è semplice: il segnale è idealmente sinusoidale, tuttavia il mondo in cui viviamo è reale, e non ideale, dunque eventuali fenomeni di distorsione all'interno del segnale (come il rumore) provocherebbero enormi errori di misura, utilizzando un tipo di voltmetro diverso da quello a valor efficace.

Data dunque una resistenza nota, un voltmetro diventa un wattmetro, semplicemente tarando la scala con grandezze quadratiche. In questo caso, l'incertezza relativa varrà:

$$\varepsilon_{P_L} = 2\varepsilon_{V_{eff}} + \varepsilon_{R_L}$$

Il discorso che abbiamo effettuato è assolutamente corretto, fatta un'ipotesi che non abbiamo finora esplicitato: il generatore di tensione o corrente del quale misuriamo la potenza, affinchè questa potenza sia un dato utile, deve essere adattato; in altre parole, l'impedenza di carico deve essere uguale al complesso coniugato dell'impedenza interna del generatore. Finora abbiamo considerato tutto adattato, ma molto spesso può capitare di non essere così fortunati: consideriamo di avere, per ipotesi, impedenze puramente resistive, e cioè non consideriamo disadattamenti sotto il punto di vista delle reattanze; al fine di avere carico adattato, la resistenza vista dal generatore dovrà essere semplicemente uguale alla sua resistenza interna.

Come possiamo garantire l'adattamento anche quando effettivamente non c'è, per poter effettuare una misura di potenza utile? Semplice! Adattando il carico, mediante un dispositivo in grado di farlo: il trasformatore. Per trasformatore intendiamo quindi semplicemente due induttori accoppiati, idealmente privi di perdite (ossia che non introducono, nel nostro circuito, errori di consumo).

L'idea fondamentale è la seguente:

Si inserisce un trasformatore con più possibilità di accoppiamenti, e in cascata un attenuatore variabile A_t, con resistenza caratteristica sempre uguale a R_0: il primo blocco serve ad adattare il generatore al resto del circuito, il secondo a variare la portata del circuito di misura.

Disponendo di più induttori, tutti accoppiati tra loro, selezionabili mediante un comando esterno, è possibile sceglierne diverse coppie, in modo da ottenere diversi rapporti di spire ingresso-uscita, considerando N_1 il numero di spire dell'induttore in ingresso (primario), ed N_2 il numero di spire dell'induttore in uscita (secondario). Supponiamo dunque di avere, a destra del secondario, una resistenza nota R_0 (propria dell'attenuatore); il generatore che stiamo studiando vedrà un'impedenza pari a:

$$R_L = R_0 \left(\frac{N_1}{N_2} \right)^2$$

La potenza attiva entrante nel trasformatore sarà pari a:

$$P_{AV} = \frac{V_E}{R_L}$$

Dove V_E è la tensione ai capi del primario del trasformatore; la potenza disponibile del generatore, ossia la potenza massima che è in grado di fornire, ad un carico adattato, si potrà calcolare facilmente a partire da un'ipotesi preliminare: consideriamo errori di consumo bassi da parte degli induttori accoppiati, e quindi potenza entrante nell'attenuatore A_t quasi uguale alla potenza attiva entrante nel trasformatore prima calcolata. Considerando V_0 la tensione ai capi dell'ingresso dell'attenuatore, la potenza in ingresso ad esso, P_0, sarà pari a:

$$P_0 = \frac{V_0^2}{R_0}$$

Considerando dunque come detto bassi errori di consumo del trasformatore:

$$P_0 \simeq P_{AV}$$

Passiamo al blocco successivo: in ingresso all'attenuatore avremo la potenza P_0 prima presentata; in uscita da esso, vedremo nuovamente una resistenza pari a R_0, nella quale entrerà una certa potenza P_M; supponendo di misurare con un misuratore M la tensione V_M ai capi di quest'ultimo resistore R_0, la potenza su di esso varrà:

$$P_M = \frac{V_M^2}{R_0}$$

Poichè consideriamo valida la relazione $P_0 \simeq P_{AV}$, possiamo dire che:

$$\frac{V_M^2}{R_0} = \frac{A_t^2 V_0^2}{R_0} = A_t^2 P_0 \simeq A_t^2 P_{AV}$$

I discorsi che abbiamo finora fatto sarebbero corretti, se fossero verificate le ipotesi sulle quali ci siamo basati; ciò purtroppo non è del tutto vero, in quanto gli induttori che vengono utilizzati per realizzare i trasformatori, per quanto di buona qualità, sono un elemento critico per il circuito che abbiamo realizzato, e senza dubbio avranno un errore di consumo non trascurabile: gli avvolgimenti con i quali sono realizzate le spire degli induttori, infatti, sicuramente dissiperanno una percentuale di potenza entrante non trascurabile. Dal momento che però le frequenze dei segnali che stiamo considerando sono relativamente basse (al limite, frequenze acustiche), possiamo pensare che gli induttori si comportino abbastanza similmente a dei corto circuiti, e quindi considerare valido tutto il nostro discorso. Si veda comunque da ciò quanto imprecise siano le misure di potenza, rispetto a quelle di tensione o corrente.

Prima di passare al secondo tipo di misure di potenza, una piccola parentesi: perchè si effettuano misure di potenza, parlando anche solo di basse frequenze? La potenza è storicamente (e non solo) il parametro più utile per indicare concetti quali efficienza, guadagno, attenuazione; caratterizzare un certo insieme di blocchi sotto il punto di vista del rapporto potenza in uscita su potenza in ingresso, significa ottenere un numero molto elevato di informazioni riguardo ad esso: abbiamo accennato al fatto che trattiamo, finora, frequenze al massimo acustiche: le tecniche che abbiamo appena introdotto, sono in grado di garantirci una misura di discreta accuratezza di potenza di amplificatori acustici, in modo da permetterci di conoscere le caratteristiche di un dispositivo piuttosto che di un altro, quali per l'appunto guadagno in potenza o efficienza.

8.2 Introduzione alle misure di potenza a radiofrequenza

Lo studio di potenze in condizioni di radiofrequenza è fondamentale, in quanto, a frequenze elevate, non è più possibile misurare una tensione, piuttosto che una corrente, in un singolo punto del circuito: si smette di parlare di circuiti a parametri concentrati, ossia circuiti dove si modellano diverse caratteristiche in un singolo dispositivo (quale una resistenza, un induttore, un condensatore); più si aumenta la frequenza, e più si evidenziano grandezze parassite, trasformando il circuito da insieme di elementi a parametri concentrati a insieme di elementi distribuiti. Parlare di misura di corrente e tensione per questo diventa impossibile, e diventa necessario parlare di misura di potenza.

Il principio sul quale si fondano le misure di potenza a radiofrequenza è molto differente da quello che siamo soliti utilizzare: la P_{RF} che misureremo ora sarà quantificata utilizzando un principio molto elementare: una potenza, dissipata su di un elemento resistivo, determina una variazione di temperatura di questo elemento: man mano che si dissipa più potenza, la temperatura dell'elemento sarà sempre più elevata; nelle nostre misure di potenza, utilizzeremo questo principio alla rovescia: misurando variazioni di temperatura $\Delta\theta$, definite come:

$$\Delta\theta = \theta - \theta_{amb}$$

Dove θ_{amb} è la temperatura ambiente. A partire da queste misure, si risalirà alla potenza dissipata sull'elemento resistivo. La misura di potenza, dunque, sarà ricondotta ad una misura di temperatura, effettuata con due strategie fondamentalmente diverse: mediante termocoppie, o bolometri.

Qualsiasi sia la tecnica che intendiamo utilizzare ai fini della misura di temperatura, alcune ipotesi devono essere assolutamente rispettate:

- Gli elementi resistivi devono essere resistenze pure: qualsiasi sia la frequenza di lavoro, nel range delle radiofrequenze, non si devono avere reattanze parassite: man mano che si alza la frequenza, infatti, tendono a risvegliarsi effetti reattivi parassiti, trascurabili a frequenze più basse; introducendo reattanze, la potenza misurata diventa lontana dalla potenza attiva, ossia quella che ci interessa misurare, dunque strumenti poco idonei sotto questo punto di vista rischiano di compromettere l'accuratezza della misura;

- A basse frequenze, la corrente si propaga in modo uniforme su di tutta la superficie del filo; alzando la frequenza, e lavorando nella fatispecie dalle radiofrequenze a onde di lunghezza inferiore alla larghezza del cavo, si manifesta un effetto molto sgradevole, detto effetto pelle, che tende a far condurre la corrente solo attraverso una corona circolare, sulla parte più esterna del cavo, per l'appunto sulla pelle; in questo modo, la resistenza offerta dai cavi è fortemente dipendente dalla frequenza del segnale. Per fare in modo da non attivare l'effetto pelle, lo spessore dei cavi deve essere inferiore alla lunghezza d'onda del segnale che inviamo, e così bisogna disporre di cavi con spessore dell'ordine del decimo di millimetro: in questo modo, la corrente sarà uniforme sia a basse frequenze che a radiofrequenza.

8.3 Sensori per la misura di temperature

Come abbiamo anticipato, esistono due filosofie di pensiero, due metodi di realizzazione di una misura di temperatura (che andrà poi ricondotta a misura di potenza): metodo utilizzante termocoppie, e metodo utilizzante bolometri. Entriamo nel merito di questi due metodi, analizzandoli nel dettaglio.

8.3.1 Termocoppie

La termocoppia è un dispositivo basato sull'effetto Seebeck: una giunzione composta da due materiali diversi, uno con una temperatura diversa dall'altro, genera una differenza di potenziale. Alla base della termocoppia vi è un chip di questo tipo:

Si mettono due giunzioni di materiali diversi, ossia due termocoppie, al fine di rendere più sensibile l'effetto Seebeck, e quindi percepire una differenza di potenziale più elevata, ai capi del voltmetro in DC; la geometria di questo circuito, di questo chip, è planare, poichè è interamente contenuta in una lamina: su due morsetti si ha l'ingresso del segnale a radiofrequenza; in questo circuito vi sono due termocoppie, formate da un resistore concentrato (realizzato mediante un film resistivo), ed un resistore diffuso, una in verso opposto all'altra (in questo modo, facendo il calcolo mediante le leggi di Kirchhoff delle tensioni, si ha una tensione risultante somma delle due). Questo circuito deve, oltre a poter creare una continua, poterla isolare e dirigere verso l'uscita, dove vi sarà un voltmetro in grado di misurare la continua; a questo scopo, si utilizzano condensatori di disaccoppiamento, ossia condensatori in grado di introdurre da una parte (a destra, sull'ingresso del voltmetro) uno zero in $f = 0$, in modo da far cadere tutta la continua ai capi del condensatore, che sarà per essa un circuito aperto, mentre per la radiofrequenza sarà un corto circuito; poichè tutta la tensione continua cade sul condensatore, la tensione rilevabile da un voltmetro attaccato in parallelo sarà la tensione Seebeck. Il condensatore sull'ingresso per la radiofrequenza (a sinistra), ha lo scopo duale: permettere solo alle radiofrequenze di entrare nel circuito, bloccando componenti continue che altererebbero la misura della tensione Seebeck.

In circuiti a radiofrequenza spesso si utilizzano cavi coassiali: questo chip termocoppia sarà dunque collegato al termine di un cavo coassiale, trasportante segnale a radiofrequenza; il coassiale ha però una geometria cilindrica, e quindi su tre dimensioni; servirà un oggetto in grado di trasformare la geometria del campo entrante nel chip da cilindrica a planare, e questo sarà formato da un contatto metallico modellato in modo da effettuare quest'operazione.

8.3.2 Sensori Bolometrici

Un bolometro è un resistore in grado di variare il proprio valore di resistenza con la temperatura; ne esistono di due categorie, classificandoli mediante il loro coefficiente termico:

- Con coefficiente termico positivo: ad un aumento di temperatura corrisponde un aumento di resistenza, e viceversa;
- Con coefficiente termico negativo: ad un aumento di temperatura corrisponde una diminuzione della resistenza, e viceversa;

Devono sempre essere verificate le ipotesi che abbiamo precedentemente proposto: questi elementi resistivi non devono essere influenzati dalla frequenza del segnale, ma solo dalla temperatura; inoltre, non devono presentare elementi di reattanza parassiti: oltre a non variare la resistenza, non deve variare l'impedenza, rimanendo sempre e comunque puramente resistiva. Sovraccarichi di segnale, inoltre, possono portare a temperature troppo elevate i bolometri, fondendoli (poichè i bolometri esaltano le variazioni di temperatura al variare della potenza in essi dissipata); un tempo i bolometri venivano realizzati semplicemente mediante fili metallici, e per questo motivo rischiavano di essere molto soggetti agli effetti sopra citati; poichè ora siamo nell'era dei semiconduttori, sono stati progettati bolometri a partire da tecnologie semiconduttive, in forma planare o in forma di piccole sferette in grado di variare sensibilmente la resistenza con la temperatura.

Abbiamo elementi in grado di variare la propria resistenza con la temperatura; al fine di poter rilevare e misurare questa differenza di resistenza, potremo utilizzare il più classico dei metodi: un ponte di resistenze (basato, come al solito, sull'idea fornita dal ponte di Wheatstone):

Il ponte, al fine di misurare la variazione di resistenza, può essere alimentato da una tensione continua, o comunque da una tensione alternata a bassa frequenza, in modo da rendere facilmente rilevabile la condizione di zero, come nella classica misura di resistenza del ponte di Wheatstone: da quest'ultimo, infatti, il ponte bolometrico non ha molte innovazioni, se non alcuni accorgimenti studiati per adattare la radiofrequenza alla tensione di alimentazione del ponte (supponiamo di inserire un solo elemento bolometrico all'interno del ponte, in modo da ricondurci facilmente allo studio classico di ponte di Wheatstone). Da un lato, come già detto il ponte deve essere alimentato da una tensione prossima ad una continua, al fine di poter rilevare la condizione di zero che ci può permettere di misurare il valore della resistenza (in questo caso, del bolometro); d'altro canto, sul bolometro deve poter scorrere la radiofrequenza, in modo da variarne la resistenza; la radiofrequenza deve scorrere solo e soltanto all'interno del resistore variabile, in modo da non venir ripartizionata per gli altri elementi del ponte, e quindi portarlo fuori bilanciamento: per questo motivo, al classico ponte di Wheatstone si deve inserire un meccanismo di disaccoppiamento, in grado di permettere ciò che abbiamo appena descritto: I_{DC} e I_{RF} devono chiudersi, di fatto, su circuiti diversi, con unico punto in comune il bolometro.

Sappiamo, dall'Elettrotecnica, che gli induttori L sono buoni conduttori per la continua (e per le frequenze basse), ma circuiti aperti per un segnale ad elevata frequenza; dualmente, le capacità C sono ottimi conduttori per le alte frequenze, ma circuiti aperti per segnali con frequenza prossima a 0.

Realizzando un circuito di questo genere, è possibile ottenere esattamente gli effetti da noi desiderati.

Montaggio pratico di sensori bolometrici

Abbiamo finora affrontato un discorso prevalentemente teorico; nella pratica, si dovrà realizzare ciò che abbiamo finora descritto. I bolometri sono oggetti molto delicati (soprattutto quelli di vecchia generazione, realizzati metallicamente); i dispositivi effettivi che verranno utilizzati come bolometri, non dovranno solo contenere questo tipo di oggetto e la schermatura prima descritta, ma anche un'interfaccia con l'esterno (con l'ingresso della radiofrequenza) idonea, adattata con il circuito in analisi, e soprattutto una protezione meccanica, e adiabatica: infatti, la misura di temperatura può essere fortemente influenzata dalla temperatura dell'ambiente esterno all'elemento bolometrico, dunque sarà necessario fornire un buon isolamento termico rispetto all'esterno.

Il connettore a radiofrequenza è in grado di non presentare effetti indesiderati anche a frequenze molto elevate; questo connettore, si adatta al tipo di cavo che si utilizza (nella fatispecie, considereremo per esempio una connessione mediante cavi coassiali). Vi sono due morsetti che si collegano al ponte, ed al resistore variabile contenuto all'interno dello scatolotto; tutto ciò protegge meccanicamente l'elemento sensibile, e lo isola termicamente verso l'esterno. Nel caso del coassiale, è sufficiente realizzare il tutto con l'anima (ossia il conduttore centrale), e la calza esterna; le terminazioni sono realizzate con un materiale resistivo.

Si mantiene quindi inalterata la geometria cilindrica del campo elettromagnetico in uscita dal filo, in modo da evitare effetti di riflessione e quindi discontinuità del campo, che provocherebbero effetti di disadattamento nel circuito. Introducendo due resistori da 100ω sui lati del circuito, in parallelo all'ingresso, avremo un'impedenza vista ai capi del cavo da 50ω, adattando quindi il bolometro con uscita coassiale ai normali sistemi a radiofrequenza. I condensatori disaccoppiano la continua, gli induttori sugli attacchi al ponte l'alternata; nella pratica, i condensatori sono realizzati mediante lamine, separate dalla calza (composta di materiale dielettrico); in questo modo, abbiamo realizzato il bolometro, interfacciabile mediante coassiale.

8.3.3 Misure di potenza mediante bolometro

Realizzato il nostro elemento bolometrico, e preparato all'utilizzo in un circuito a ponte, possiamo iniziare a parlare di misura di potenza dissipata, con segnali a radiofrequenza. Il problema principale, è determinare un metodo in grado di far mantenere costante, negli stati di regime (ossia una volta terminati i transitori), il valore della resistenza del bolometro: l'introduzione del volometro non deve in alcun modo poter disadattare il carico. Per far ciò, seguiamo il seguente ragionamento: considerando sempre la misura effettuata con il ponte in stato di equilibrio, il bolometro è percorso da due correnti: una continua, ed una a radiofrequenza. Questo resistore, quindi, si scalderà sia per via della continua, che per via della radiofrequenza: il nostro scopo sarà distinguere le due cose, comprendendo quale contributo deriva da quale dei segnali. Innanzitutto, noi sappiamo che la variazione di temperatura del bolometro è legato alla legge:

$$\Delta\theta = \frac{P_{diss}}{K_\theta}$$

Dove θ è la temperatura in cui ci si trova, e K_θ il coefficiente termico del bolometro. Per aver adattamento del circuito, serve resistenza costante, ma dunque anche temperatura θ costante, in modo da poter dir di avere sempre la stessa potenza dissipata sull'elemento resistivo. Per questo motivo, si distinguono due fasi nel processo di misurazione:

1. Si alimenta il ponte, senza inserire il contributo del segnale a radiofrequenza: $P_{RF} = 0$. In questo modo, il ponte è alimentato esclusivamente dalla continua; una volta portato in condizione di equilibrio, non vi sarà corrente nella diagonale, e sarà possibile quantificare la potenza dissipata dal bolometro come:

   $$P_{DC,1} = R_0 \left(\frac{I_1}{2} \right)^2$$

   Dove R_0 è uguale alla resistenza del bolometro, I_1 è la corrente che esce dal generatore, e viene ripartita in due metà identiche tra i due rami del ponte (poichè si hanno resistenze equivalenti uguali, per semplicità).

2. Si introduce la radiofrequenza nel circuito, ed il ponte andrà fuori equilibrio, poichè l'introduzione della radiofrequenza aumenta (o diminuisce, a seconda del coefficiente termico del bolometro), la resistenza opposta da R_{bol}.

   Al fine di riportare il ponte in condizione di equilibrio, modifichiamo il valore della resistenza di generatore, R_G, dissipando una parte ulteriore della continua, e riportando quindi il ponte in condizione di equilibrio. Dalla prima fase possiamo ricavare la potenza dissipata dalla sola corrente continua, $P_{DC,1}$; da questa seconda fase, modificato il valore di R_G, possiamo ricavare la potenza dissipata dalla sola continua, con anche la radiofrequenza nel bolometro, ottenendo:

   $$P_{DC,2} + P_{RF} = R_0 \left(\frac{I_2}{2} \right)^2$$

Date queste relazioni, la potenza del segnale dovuto al solo contributo di radiofrequenza, sarà semplicemente la differenza delle due:

$$P_{RF} = R_0 \left[\left(\frac{I_1}{2} \right)^2 - \left(\frac{I_2}{2} \right)^2 \right]$$

8.3.4 Effetti termici, disadattamenti e perdite

Come ciascuno dei metodi di misura che abbiamo analizzato, anche questo contiene molte imperfezioni, dovute al mondo reale in cui viviamo; possiamo tuttavia cercar di migliorare i nostri sistemi di misura, al fine di renderli sempre più accurati, e quindi in grado di fornirci una stima sempre più realistica del valore. Considereremo due tra le principali fonti di indeterminazione, cercando di spiegare come attenuare i problemi da esse derivanti.

Instabilità della temperatura ambiente

Le misure di potenza che effettuiamo si basano sulla misura di una variazione della resistenza rispetto ad un campione che dovrebbe essere fisso e ben definito, ossia la temperatura ambiente. Come ben sappiamo, tuttavia, la temperatura ambiente è tutt'altro che stabile e ben definita: la misura della variazione rispetto alla temperatura ambiente potrebbe dunque presentare una forte indeterminazione, derivante dal fatto che il campione di riferimento non è per niente definito. Si può ovviare in maniera relativamente facile ad un problema di questo tipo, utilizzando una piccola astuzia: una seconda termocoppia, collegata ad un secondo ponte, che misurerà sempre la temperatura ambiente; questo ponte avrà un sistema in grado di termostatizzare la temperatura, in modo da mantenerla costante rispetto a variazioni non troppo elevate della temperatura ambiente, desensibilizzando il primo ponte rispetto ai transitori. La variazione di temperatura $\Delta\theta$ sarà considerata in riferimento non più alla temperatura ambiente realmente misurata, ma rispetto al ponte di compensazione appena introdotto.

Disadattamenti e perdite dovute alla radiofrequenza

Abbiamo parlato dell'interfaccia del ponte bolometrico alla radiofrequenza in modo un po' troppo ottimistico: la nostra idea, in effetti, era quella di misurare la potenza attiva disponibile del generatore di tensione E, polarizzante con una continua il ponte, come:

$$P_{AV} = \frac{E^2}{4R_0}$$

Poichè disponiamo di resistenze tutte pari a R_0; dalle teorie sui Campi Elettromagnetici, tuttavia, sappiamo che l'interfaccia del bolometro verso la radiofrequenza, provoca una perdita per riflessione di parte del campo, e dunque un dissipamento di potenza; si ha dunque una riflessione, che comporta alla dispersione di una parte della potenza per riflessione, parte che chiameremo P_R; la potenza attiva P_{AV} emessa dal generatore e la potenza in ingresso al bolometro, P_{IN}, non coincideranno, poichè:

$$P_{IN} = P_{AV} - P_R$$

Oltretutto, la porzione di potenza riflessa P_R dipende dalla frequenza del segnale a radiofrequenza che inseriamo: ciò introduce ulteriori errori ed ulteriori indeterminazioni nel nostro sistema.

Supponendo che la radiofrequenza non eliminasse l'adattamento, o meglio che non vi sia l'effetto elettromagnetico di riflessione, si aggiungerebbe un altro problema: una potenza P_D dissipata sulle pareti del montaggio bolometrico; considerando l'ipotesi che $P_R = 0$, dunque, la potenza effettivamente misurata, P_{MIS}, sarebbe uguale a:

$$P_{MIS} = P_{IN} - P_D$$

La soluzione ai problemi intrinseci del bolometro (riflessione e dispersione) non consiste tanto nell'operare fisicamente sul bolometro in modo da migliorarlo in qualche maniera, ma solo quella di saper quantificare l'errore, mediante un processo di taratura del resistore variabile considerante sia gli effetti di P_R che gli effetti di P_D; il costruttore dovrà dichiarare, al variare della frequenza f, i fattori di correzione con i quali migliorare la misura, a partire dal valore misurato P_{MIS}.

Capitolo 9

Generatori di Segnali

Realizzare generatori di segnali può essere molto importante, sotto molti punti di vista: mediante essi è possibile effettuare test di vario genere, su apparecchiature audio, radio, ad alta frequenza o meno, o anche su sistemi di controllo di diversa natura; particolarmente importanti tra questi, sono i generatori di segnali sinusoidali, a partire dai quali (e non solo) si vedrà poi come realizzare sistemi più complessi, in grado di generare forme d'onda molto più variegate (quali triangoli, onde quadre, e simili).

9.1 Generatori sinusoidali

Come accennato, i generatori sinusoidali sono i più interessanti tra tutti i generatori di segnali esistenti. Le caratteristiche dei segnali sinusoidali che dobbiamo essere in grado di modificare, in generatori di segnali di questo tipo, sono sostanzialmente frequenza (possibilmente, permettendo una variazione tendenzialmente continua su di un range più ampio possibile), e regolazione dell'ampiezza. Mediante ulteriori comandi sarà possibile modificare anche parametri quali la fase, e altri. Oltre a permettere queste operazioni, si richiede che le grandezze sintetizzate siano ben definite: si vuole avere una buona stabilità di frequenza, ossia una scarsa variazione nel tempo della frequenza, ed una bassa distorsione di armoniche, ossia non si vuole aver troppo rumore introdotto da armoniche secondarie.

Anche a seconda del range di frequenze che si intende rappresentare, introducendo la solita distinzione tra basse frequenze (BF), e radiofrequenze (RF), si utilizzeranno tecniche diverse (e schemi diversi, ovviamente), per la costruzione di dispositivi di questo tipo.

Come si realizza, in linea di massima, un generatore sinusoidale? Alla base di tutto vi è un oscillatore sinusoidale, ossia un dispositivo in grado di effettuare un'oscillazione, per l'appunto avente forma sinusoidale. Questo deve essere, in qualche maniera, regolabile mediante controlli esterni; l'output di questo viene mandato in ingresso ad un amplificatore separatore, ossia un dispositivo in grado di creare alta impedenza di carico per il generatore sinusoidale, in modo da lasciarlo oscillare tranquillamente, senza fargli sentire l'influenza di eventi ad esso esterni. L'uscita del separatore viene amplificata da un amplificatore di potenza, la cui uscita verrà poi regolata da un attenuatore a scatti (tarato). Possiamo ritenere così schematizzato il caso più generale di generatore sinusoidale, senza essere entrati nel merito dei singoli elementi che lo costituiscono; dal momento che parlare di generatori a bassa frequenza ed ad alta frequenza richiede metodi molto diversi, si parlerà nei dettagli solo una volta stabilito il range di frequenze che si intende trattare.

9.1.1 Generatori sinusoidali a bassa frequenza

I generatori sinusoidali a bassa frequenza possono essere molto utili in molti contesti, specialmente nel campo dell'elettronica: dal momento che le frequenze dell'acustica sono da considerarsi nel range delle frequenze basse, possiamo intuire che i generatori sinusoidali ci permetteranno di caratterizzare dispositivi di vario genere. Generatori basati su questo principio fondamentalmente hanno frequenze variabili da qualche decina di Hz a qualche centinaio di kHz; solitamente, l'impedenza di uscita di dispositivi di questo genere sarà o 50 Ω o 600 Ω; l'accuratezza di taratura di frequenza, invece, avrà un ordine di grandezza del tipo:

$$\frac{\frac{\delta f}{f}}{\Delta t} \sim \frac{10^{-3}}{h}$$

Si ha una distorsione causata da armoniche successive circa pari al 3%, tensioni di uscita dell'ordine di $0,1 \div 10$ V, e correnti massime di $0,1$ A.

Alla base dei generatori sinusoidali a bassa frequenza, vi è sostanzialmente un doppio sistema di retroazione, una reazione, ossia un feedback positivo, ed una controreazione, ossia un feedback negativo.

- La rete di reazione positiva è costituita dalla rete selettiva: essa sceglie una certa frequenza, e la manda all'amplificatore operazionale, continuando ad incrementarla;
- La rete di controreazione sarà la rete di controllo di ampiezza: se la rete selettiva continuerà a selezionare sempre la stessa frequenza, e l'operazionale ad ampliarla, capiterà che, alla lunga, l'oscillazione raggiunga ampiezze spropositate; la controreazione stabilizza l'ampiezza delle oscillazioni, vincolandola ad un certo valore massimo.

Lo schema di principio ricorderebbe in realtà un circuito a ponte (detto ponte di Wien):

Cerchiamo di capire come potrebbe fungere questa retroazione: effettuando un taglio del circuito, ed introducendo qui un generatore di segnale sinusoidale, questo viene amplificato e mandato indietro dal blocco di retroazione β, che lo riduce. Il blocco di guadagno A dunque amplifica il segnale (senza invertirlo), il β lo retroaziona, rimandandolo indietro. Modificando i valori delle resistenze, si modifica il guadagno del sistema amplificatore A: di fatto l'operazionale ha un'impedenza elevatissima (idealmente infinita) in ingresso, poichè $A \to \infty$ (trattandosi di un operazionale); possiamo dunque schematizzare il comportamento delle reti, caratterizzando i blocchi A e β, calcolando il guadagno di anello separando i due componenti:

$$A = \frac{V_{out}}{V_+} = \frac{R_3 + R_4}{R_4}$$

$$\beta = \frac{V_2}{V_{out}} = \frac{Z_2}{Z_1 + Z_2}$$

Dove identifichiamo i vari parametri qua presentati come:

$$Z_1 = R_1 + \frac{1}{sC_1}; \quad Z_2 = \frac{R_2}{1 + sR_2C_2}$$

Come vediamo, il guadagno A è una quantità puramente reale: esso dunque provocherà una variazione di ampiezza del segnale, ma non ne varierà la fase. Il fattore β è una quantità complessa, e dunque al variare della frequenza presenterà comportamenti diversi. Un caso molto interessante e molto particolare, è quello tale per cui $R_1 = R_2$, e $C_1 = C_2$: in questo caso, il circuito risulta essere di fatto un circuito risonante: si avrà dunque un punto di pulsazione a fase 0, ω_0, ossia la condizione di risonanza del circuito.

Il valore ω_0 si può calcolare semplicemente come:

$$\omega_0 = \frac{1}{RC}$$

Dove $C = C_1 = C_2$ e $R = R_1 = R_2$.

Essendo la fase della rete 0, e quindi il blocco di retroazione β ha un contributo quantificabile: si ha infatti un massimo assoluto della funzione di trasferimento, calcolabile mediante le formule prima proposte, ottenendo quindi:

$$\beta_{max} = \left| \frac{V_2(\omega)}{V_1(\omega)} \right|_{\omega=\omega_0} = \frac{1}{3}$$

A questo punto, si può notare che, se scegliamo un segnale V_1 in ingresso, in uscita avremo il medesimo, normalizzato di 3:

$$V_1 \longrightarrow V_2 = \frac{V_1}{3}$$

Potrebbe dunque addirittura capitare di avere un guadagno di anello positivo, e dunque una retroazione sostanzialmente positiva (ossia di non avere l'effetto di ridimensionamento): se capita infatti che:

$$|A\beta| > 1 \Longleftrightarrow |A| > 3$$

Allora capita che si ha un effetto complessivo di rigenerazione del segnale: si ha un minimo guadagno di anello, dunque la retroazione negativa (la controreazione) non è in grado di ridimensionare il segnale, che continuerà a crescere.

Ricordiamo che, ora come ora, abbiamo ancora l'anello aperto: togliamo il generatore sinusoidale di prova, chiudiamo l'anello, e cerchiamo di capire cosa capiterà: abbiamo un guadagno di anello sostanzialmente positivo, ma nessun segnale da amplificare introdotto. Il fatto che non ci si trovi a temperature pari a 0 K, ci fa intuire che ci sarà di sicuro del rumore nella rete di retroazione, quantomeno termico. Dal momento che la funzione di trasferimento di questo oggetto è sostanzialmente un amplificatore passa banda, fissato a frequenza ω_0 (ossia la frequenza di risonanza del circuito che abbiamo considerato, tutte le componenti del rumore verranno tagliate, tranne una. Questa andrà nell'operazionale, dove verrà amplificata, e così via. In questo modo, dal momento che solo una componente spettrale è sostanzialmente amplificata, avremo in uscita da questa rete una sinusoide.

Questo procedimento si può ricordare in un caso molto comune a chi suona, o meglio cnata: avvicinando un microfono ad una cassa, si ha un guadagno di anello positivo: il rumore che c'è in aria entra nel microfono, viene amplificato dall'amplificatore che lo fa uscire dalla cassa, da qui rientra nel microfono, viene riamplificato, e si sente un fischio, dopo qualche istante: questo è la sinusoide amplificata mediante la retroazione creata dalla chiusura dell'anello, realizzata mediante l'avvicinamento del microfono alla cassa.

Possiamo capire che dunque l'armonica derivi semplicemente da una singola componente del rumore, che viene continuamente amplificata dalla rete di reazione, usando le condizioni di innesco prima espresse (ossia, $|A| > 3$). Al fine di avere una sinusoide ben definita, serve che la funzione di trasferimento della rete sia molto schiacciata, ossia sia molto stretta: il passa banda deve essere estremamente selettivo, e quindi avere una banda passante molto ridotta. Questo si ottiene ottimizzando il parametro Q del circuito: un Q

elevato implica un massimo della funzione di trasferimento molto elevato, e se questo è elevato significa che la curva è molto stretta.

Per poi limitare il continuo crescere in ampiezza di questa sinusoide, vi sono sostanzialmente due tecniche: una intrinseca al circuito, o meglio alle caratteristiche dell'amplificatore operazionale: poichè quest'ultimo dopo un certo tempo raggiunge una condizione di saturazione, il sistema tende ad andare sotto la condizione di innesco e così a bloccarsi nel crescere; esiste un sistema di controllo più automatizzato (ed affidabile) rispetto a questo trucco, basato sulle caratteristiche dell'amplificatore operazionale: considerando lo schema finora discusso, introducendo al posto dei resistori collegati al morsetto $-$ dell'operazionale dei bolometri, ossia delle resistenze variabili con la temperatura (elementi non lineari), con l'aumentare dell'ampiezza, aumenta anche la potenza in essi dissipata; se il bolometro R_3 è un termistore, ossia a coefficiente termico negativo, e cioè diminuisce la propria resistenza all'aumentare della tensione su di esso dissipata; dualmente, R_4 sarà un resistore metallico, ossia a coefficiente termico positivo (aumento temperatura, aumento resistenza). In questo modo verrà eliminata sempre più tensione man mano che uscirà una tensione elevata dall'operazionale, autoregolando la tensione. Utilizzando questo tipo di aggiustamento, l'ampiezza delle oscillazioni risulta essere molto ben stabilizzata; inoltre, l'amplificatore lavora in zona lineare, e dunque non si hanno molti contributi di distorsione di armoniche secondarie.

Come abbiamo detto, questo tipo di dispositivi per la generazione di frequenza, ha un range di frequenze generabili da 10 Hz fino a 100 kHz (circa, come ordine di grandezza); vediamo perchè:

- Per quanto riguarda il limite inferiore di frequenza, dal momento che la costante di tempo termica delle resistenze è molto bassa, e dal momento che il segnale è una grandezza variabile nel tempo, il guadagno del sistema retroazionato tende a fluttuare con le resistenze; le resistenze della rete di Wien, inoltre, portano enormi errori di consumo, introducendo una forte non idealità;

- Per quanto riguarda il limite superiore di frequenza, dal momento che l'amplificatore operazionale non è un elemento reale, si attivano le capacità parassite (i cui poli sono stati peraltro spostati molto avanti grazie all'anello di retroazione), e così il guadagno dell'amplificatore inizia a crollare; a questo punto l'operazionale inizia a fare una partizione con il ponte di Wien, dal momento che aumenta l'impedenza in uscita, eliminando così la funzionalità del circuito.

Come si regola la frequenza, per farla variare (ovviamente restando nel range di validità)? Mediante la modifica della frequenza (o pulsazione) di risonanza del circuito, f_0 (o ω_0), definita come:

$$f_0 = \frac{1}{2\pi RC}$$

Si può dunque agire sia sui resistori che sulle capacità; unico accorgimento nel farlo, è variare contemporaneamente tutte le resistenze e/o tutte le capacità della rete di Wien, in modo da mantenere intatte le caratteristiche finora ricavate dal circuito, e poter considerare validi tutti i discorsi finora detti.

Come attenuatore di uscita, ci son diversi tipi di circuiti:

- Nell'ambito di generatori di bassa qualità, è possibile utilizzare un banalissimo potenziometro: in questo ambito si possono sì effettuare variazioni, ma non tarare la tensione di uscita. A seconda dell'ampiezza del segnale, l'impedenza di uscita inoltre subisce variazioni, quindi non si tratta proprio di un sistema affidabile.

- Esistono, nell'ambito di generatori di qualità superiore, reti attenuatrici ad attenuazione tarata, e resistenza caratteristica costante con la frequenza e con l'attenuazione.

  Considerando R_L la resistenza di carico, la tensione sul carico V_L si può misurare misurando la tensione in un'altra posizione del circuito, nella fatispecie V_m, conoscendo il fattore di attenuazione K (selezionato da noi), e ponendo che vi sia un partitore $\frac{1}{2} - \frac{1}{2}$, tra R_L e R_0: per far questo, $R_L = R_0$. In questo caso, infatti:

  $$V_L = \frac{K}{2} V_m$$

  Si noti che, se il carico non è adattato ($R_L \gg R_0$), la tensione di uscita è tendenzialmente il doppio di quella che si può misurare con l'indicatore, col misuratore di tensione in uscita dal generatore V_0.

9.1.2 Generatori sinusoidali a radiofrequenza

Concettualmente, i generatori sinusoidali a radiofrequenza non sono così differenti da tutto ciò che abbiamo finora utilizzato; bisognerà però avere alcuni accorgimenti, che prima non era necessario prendere (dal momento che si utilizzavano basse frequenze), come si descriverà tra breve.

Generatori a radiofrequenza (RF) possono essere utilizzati per rilevare le caratteristiche dei bipoli o doppi bipoli in situazioni per l'appunto di segnali a frequenza elevata: con i metodi che descriveremo, infatti, si avrà una gamma di frequenze variabile dai 500 MHz al GHz circa, come ordine di grandezza; in fatto di deriva di frequenza, ossia variazione della frequenza in un intervallo di tempo Δt, si ha una buona stabilità:

$$\frac{\frac{\delta f}{f}}{\Delta t} \sim 10^{-4}$$

L'accuratezza della taratura delle frequenze è buona: dal 1% al 3%; la distorsione di armoniche successive di solito ha un contributo variante da -30 dB fino a -40 dB rispetto alla fondamentale; si ha inoltre possibilità di effettuare modulazione analogiche di vario genere (AM, FM, PM) sul segnale generato. Le caratteristiche non sono dunque così proibitive come si potrebbe pensare pensando alla radiofrequenza, anche se non abbiamo ancora discusso i problemi reali legati a dispositivi di questo tipo.

Parlare di radiofrequenza significa parlare di effetti parassiti: se prima ci è stato infatti possibile realizzare un circuito a ponte di Wien, basato sostanzialmente sull'utilizzo di resistori, ora non ci sarà assolutamente possibile, dal momento che i resistori sono elementi fortemente instabili al variare della frequenza, quando ci si trova nel range delle radiofrequenze, e quindi sarà necessario utilizzare qualche artificio di tipo diverso per realizzare il circuito.

Si realizzano nella fatispecie oscillatori di due tipi: oscillatori Hartley (a partizione induttiva), od oscillatori Colpitts (a partizione capacitiva).

La frequenza del dispositivo potrà essere variata semplicemente mediante una reattanza variabile; usando nella fatispecie un condensatore variabile, potremo modificare il parametro frequenza a nostra discrezione (ovviamente secondo i limiti dettati dalle incertezze introdotte dall'elemento variabile, e quelli del suo range di variazione).

Abbiamo un oscillatore a radiofrequenza, che viene collegato ad un amplificatore separatore, ossia ad un disaccoppiatore, unito ad un amplificatore di potenza. Fuori vi è un modulatore AM o FM (a seconda del tipo di dispositivo), utilizzante un oscillatore a bassa frequenza interna per effettuare l'operazione di modulazione. Una retroazione realizza la stabilizzazione della frequenza, ed un attenuatore tarato variabile permette di regolare l'output del generatore di segnale.

Poichè si parla di RF, si parla automaticamente anche di schermatura: utilizzare frequenze elevate è molto complicato, poichè bisogna evitare di subire perdite o variazioni di informazioni; per schermare bisogna considerare in particolare due tipi di elementi da cui è necessario schermarsi:

- L'ambiente esterno al circuito;
- Gli altri elementi elettronici interni al circuito costituente il generatore.

L'ambiente esterno non deve influire, nel senso che la radiofrequenza potrebbe venir irradiata dal cordone di alimentazione del circuito, e/o da altri punti della struttura; schermando in maniera opportuna, dunque, si potrà eliminare la possibilità del segnale a radiofrequenza di uscire da posti diversi del punto V_{out}, ossia la direzione che deve prendere nel circuito. Di fatto anche gli altri componenti possono essere abbastanza influenzabili tra di loro (e dall'esterno), quindi sarà necessario schermare, oltre a tutto il circuito dall'esterno, anche i singoli elementi tra loro.

Schermare in questo ambito significa semplicemente chiudere in una scatola metallica ciascuno dei componenti influenzabili, definendo esclusivamente i punti di uscita, i percorsi del segnale, tra una scatola e l'altra, e chiudere tutto il circuito in un'unica scatola, ponendo ciascuno degli schermi allo stesso potenziale di riferimento (ottenendo quindi di fatto una superficie equipotenziale); la scatola, la carcassa, sarà di solito il potenziale di riferimento in questione.

I generatori realizzati mediante un adattamento in circuito LC dei vecchi, presentano diversi inconvenienti: uno tra tutti è lo scarso range di variazione (al massimo pari ad un paio di ottave, parlando in termini logaritmici). Il segnale inoltre, a causa delle tecniche utilizzate per la realizzazione di questo tipo di generatore, spesso capita che lo stesso segnale, in range di frequenza diversi, presenti ampiezze diverse. L'unica soluzione per problemi di questo tipo, situati alla radice del circuito utilizzato, van risolti cambiando circuito: utilizzando tecniche ed idee completamente diverse, vedremo che si potrà far di meglio di ciò che abbiamo finora ideato.

9.1.3 Generatori a battimenti

Il generatore di battimenti si basa su un'idea completamente differente da quella finora utilizzata: al posto di una singola frequenza ottenuta mediante agganciamento, si utilizzerà una frequenza sintetizzata mediante combinazione di due oscillazioni.

Basiamo il tutto sul nostro elemento precedente, ossia l'oscillatore LC; le due sinusoidi, sintetizzate dunque mediante gli oscillatori già descritti in precedenza, vengono introdotte in ingresso ad un mescolatore (ossia ad un mixer). Il mixer, in uscita, fornisce l'insieme delle combinazioni lineari di tutte le armoniche introdotte al suo interno. Si avrà dunque uno spettro formato da un certo numero di righe, ciascuna rappresentante una delle combinazioni lineari:

$$|\alpha f_1 \pm \beta f_2|$$

Di tutte queste righe, ce ne interessa una in particolare: la differenza delle due frequenze introdotte dal mixer, ossia:

$$|f_2 - f_1|$$

Per selezionare questa riga, basterà utilizzare un filtro passa banda (o spesso anche solo un semplice passa basso).

La cosa interessante di questo sistema, è il fatto che, facendo variare una sola delle frequenze (per esempio consideriamo variabile solo f_2), l'uscita del sistema varierà notevolmente. Il risultato notevole, è il fatto che è possibile variare, in termini relativi, la frequenza uscente dal sistema mixer + filtro, in modo enorme, ossia con un range molto, molto più elevato rispetto a quello di variazione di f_2. L'uscita f_u, nella fatispecie, varierà tra:

$$f_u \in [f_{2,min} - f_1; f_{2,max} - f_1]$$

Scegliendo opportunamente le frequenze f_1 e f_2, ossia scegliendole ad una frequenza molto più elevata rispetto al range che vogliamo rappresentare, e vicine tra loro, possiamo ottenere dunque una rappresentazione molto più elevata rispetto a quella che potremmo usare variando direttamente le frequenze. Supponiamo, in un esempio pratico, di avere f_1 = 10 MHz; f_2 variabile da 10,001 a 12 MHz, con continuità. Poichè si ha che:

$$f_u = |f_2 - f_1|$$

Si ha che f_1 può variare da 1 kHz a 2 MHz, raggiungendo dunque oltre 3 decadi di distanza tra la frequenza minima e la frequenza massima ottenibili (contro le 2 ottave rappresentabili dal sistema precedente). Questo, con una variazione relativa delle due del solo 20%.

Vi è un rovescio della medaglia: il fatto che si abbia una frequenza definita come differenza di due frequenze (di fatto senza correlazione tra loro), genera una notevole incertezza sulla frequenza equivalente. Utilizzando la propagazione delle incertezze, infatti:

$$\Delta f_u = \Delta f_2 - \Delta f_1$$

Considerando variazioni relative:

$$\frac{\Delta f_u}{f_u} = \frac{\Delta f_2}{f_u} - \frac{\Delta f_1}{f_u} = \frac{\Delta f_2}{f_2 - f_1} - \frac{\Delta f_1}{f_2 - f_1} =$$

$$= \frac{\Delta f_2}{f_2}\frac{f_2}{f_2 - f_1} - \frac{\Delta f_1}{f_1}\frac{f_1}{f_2 - f_1}$$

Per come abbiamo scelto le frequenze f_1 e f_2, ossia molto elevate rispetto al loro range di variazione, avremo che le frazioni affiancate alle variazioni relative saranno molto elevate, e di conseguenza lo sarà la variazione relativa della frequenza di uscita f_u. Variazione relativa molto elevata significa che per una minima perturbazione si sente una perturbazione molto elevata dell'uscita, cosa che potrebbe dunque amplificare i contributi delle incertezze. Le variazioni qui introdotte, sono dovute ad un mix di contributi: l'instabilità della frequenza nel tempo, l'instabilità della frequenza al variare della temperatura, e l'incertezza sulla determinazione delle frequenze.

L'unico modo per abbassare quest'enorme variazione di frequenza, è ridurre i termini moltiplicanti gli indicatori delle variazioni relative, e per far ciò bisogna porre dei limiti alle minime frequenze che si possono ottenere in uscita. Si può tentare di compensare le derive di frequenze dovute ad effetti termici, se queste sono deterministiche: si può considerare queste incertezze (ed eventualmente quelle di deriva nel tempo, se fossero deterministiche anch'esse), come incertezze di categoria B, e dunque cercare di limitarle utilizzando alcuni accorgimenti teorici a posteriori. Nel caso di grandezze non deterministiche, non si può fare niente, e dunque, utilizzando per forza un modello di worst case, sarà necessario sommare in modulo tutte le grandezze.

VCO: Voltage-Controlled Oscillator

I generatori a battimenti si prestano molto bene per la realizzazione di generatori sweep, ossia generatori di segnali sinusoidali in grado di variare la propria frequenza con continuità; in particolar modo, sono interessanti i generatori VCO, ossia quelli in grado di variare la propria frequenza di oscillazione mediante un controllo in tensione. Una tensione dunque comanda una scansione di frequenze, facendo variare per l'appunto con uno sweep, ossia una variazione continua, la frequenza dell'oscillatore. Il range di variazione è piuttosto elevato, poichè il principio alla base di tutto è un generatore a battimenti, e dunque è possibile utilizzare questo tipo di generatori per diverse applicazioni, quali la modulazione di ampiezza o frequenza, e la misura di caratteristiche di bipoli/doppi bipoli.

9.2 Generatori di forme d'onda

I generatori di forme d'onda, o generatori di funzioni, rappresentano la naturale estensione di quelli che abbiamo finora trattato, ossia dei generatori sinusoidali. Con essi si possono rappresentare diverse forme d'onda: oltre alle classiche sinusoidi, si possono rappresentare infatti triangoli, onde quadre, rampe, impulsi a duty cycle variabile generici, con frequenze variabili da valori molto bassi (anche fino a 10^{-3} Hz) a decine di MHz. La caratteristica di poter dar vita a segnali la cui frequenza li rende molto simili ad una continua, permette di studiare sistemi con costanti di tempo molto elevate, ossia con una forte inerzia; spesso, inoltre, vengono sostituiti ai normali generatori a ponte di Wien, anche nell'ambito dei loro utilizzi più classici (quali quelli audio). Spesso rimpiazzano anche i generatori a battimenti, per quanto riguarda i generatori sweep.

Si utilizza uno schema molto differente da quelli finora studiati, come descriveremo tra breve.

Si usa fondamentalmente un generatore di rampe, basato su di un commutatore tra due sorgenti di corrente, I_1 e I_2. Per un certo tempo t_1 vi sarà attivo I_1, per t_2, I_2. Commutando i due generatori di corrente, sarà possibile variare il segnale in ingresso al circuito; il resto del sistema è legato a comparatori di soglia: ogni rampa generata dalle commutazioni viene continuamente regolata da un comparatore di soglia; quando la rampa raggiunge una certa ampiezza, si ha una commutazione; da qui, si possono effettuare operazioni di vario tipo, quali integrazione (formando triangoli), o trasformazioni, mediante circuiti di vario tipo, in forme d'onda di altro genere, quali onde quadre o sinusoidi. Modificare la forma d'onda triangolare è semplice: variando il rapporto delle due correnti, I_1 e I_2, si possono ottenere effetti di variazione di pendenza dei triangoli, o distanza picco-picco nelle onde quadre, o ampiezza delle sinusoidi. Per variare il duty cycle, il punto centrale delle forme d'onda, è sufficiente lavorare sui t: modificando t_1 e t_2, si modifica il duty cycle della forma d'onda.

Disponendo di un circuito generatore di triangoli, si può operare su di essi dunque semplicemente introducendo ulteriori circuiti, in grado di effettuare particolari operazioni. Particolarmente interessante può essere formare, a partire da un triangolo, un segnale di tipo sinusoidale: ciò si può ottenere ad esempio introducendo un bolometro, ossia un resistore in grado di variare la propria resistenza con la temperatura (o fare la stessa cosa con un resistore comandabile in tensione, per esempio); altro modo di realizzare una sinusoide a partire da un triangolo è introdurre un circuito dotato di diversi istanti di commutazione, modellante il proprio guadagno ad ogni commutazione. Abbiamo visto in questo modo alcuni esempi di utilizzo dei segnali triangolari, prodotti dal circuito appena presentato.

9.2.1 Oscillatori al quarzo

L'elemento alla base di strumenti di questo tipo, è l'oscillatore; esso viene solitamente realizzato mediante quarzi, naturali o artificiali (riprodurre quarzi in laboratorio è infatti un'operazione assolutamente realizzabile). Il quarzo è un elemento piezoelettrico, ossia il grado di produrre una corrente elettrica se sottoposto a dilatazioni o compressioni meccaniche. Per sfruttare questa cosa, si taglia il cristallo di quarzo in piastrine, seguendo gli assi cristallografici del materiale. Queste vengono metallizzate, ossia viene introdotto un contatto ohmico ai loro lati, e si introducono in una sorta di condensatore, utilizzante il quarzo come dielettrico. Si sfrutta dunque l'effetto piezoelettrico, in maniera leggermente diversa da quella descritta: si fa scorrere nel condensatore corrente, sfruttando la reversibilità del processo (introducendo una corrente, infatti, il quarzo si comprime; la compressione introduce una corrente elettrica, e così via fino a quando scorre corrente nei contatti ohmici del quarzo); il quarzo si comporta dunque al contempo come carico, e come generatore di corrente. L'idea di utilizzarlo come dielettrico per un condensatore inoltre è azzeccata poichè esso è un buon dielettrico, presenta un'elevata resistenza a bassa frequenza, e si modella dunque come una reattanza capacitiva. Aumentando la frequenza del segnale, si raggiunge una condizione confrontabile alla risonanza: da una certa frequenza in su, infatti, il materiale piezoelettrico smette di comportarsi come una capacità, iniziando ad opporre una notevole impedenza anche alle frequenze elevate, comportandosi dunque in maniera induttiva. Aumentando ulteriormente la frequenza, dunque, si ha una discontinuità, che riporta il quarzo ad un comportamento capacitivo, perdendo resistenza rispetto alla frequenza, introducendo una nuova risonanza.

Il quarzo si può dunque modellare con un circuito RLC serie, in parallelo ad un condensatore: a frequenze basse, si ha un comportamento prevalentemente capacitivo, che poi, in seguito ad una certa frequenza di risonanza, commuta in un comportamento induttivo. Da questa frequenza in su, si può pensar di avere sostanzialmente a che fare dunque con un induttore. Poichè in un certo punto sull'asse delle frequenze si ha un polo, una seconda frequenza di risonanza, si può pensare di avere un condensatore in parallelo a questo induttore, riportante il quarzo ad un comportamento prevalentemente capacitivo. La resistenza serie, comprendente le perdite di tipo elettrico/meccanico del quarzo, è molto bassa, e dunque si può pensare che il fattore di qualità Q del circuito risonante sia molto elevato. Non introducendo troppi stimoli meccanici ed elettrici sul quarzo, dunque, si può evitare di introdurre fattori di perdita, e quindi mantenere un fattore di qualità vicino a quello intrinsico.

9.2.2 Oscillatore

L'oscillatore si basa su di uno schema Colpitts, in cui il piezorisonatore vien fatto lavorare alla frequenza in cui ha comportamento prevalentemente induttivo. Mediante un termostato, si stabilizza il punto di lavoro del quarzo, evitando derive termiche. Il quarzo in questo modo ha una frequenza elevata, molto stabile, ma fissa: essa è ritoccabile mediante l'introduzione di capacità, che modificano il comportamento circuitale con il quale abbiamo modellizzato il quarzo, ma si tratta di ritocchi, non di variazioni in grado di realizzare un generatore di segnali. Una volta termostatizzati, i quarzi presentano una stabilità notevole, ma comunque non assoluta: al variare del tempo t, infatti, vi è un aumento della frequenza dell'oscillatore.

Si ha, dal momento dell'accensione, un primo istante di assestamento, dopo di che, raggiunto uno stato di termostatizzazione e di regime, si ha una costante crescita della frequenza, lineare, con pendenza α. Si ha dunque la possibilità di tarare questo strumento con una certa accuratezza, e conoscere il valore della frequenza iniziale sapendo da quanto tempo è stato tarato lo strumento: poichè la legge è lineare, infatti, si può dire che in un tempo t_1, rispetto al tempo di taratura a regime t_0, si ha:

$$f(t_1) = f(t_0) + \alpha(t_1 - t_0)$$

Una nota importante: l'oscillatore non si può mai spegnere; infatti, se spegnessimo l'oscillatore, perderemmo la taratura, poichè il regime riassesterebbe sì ad un certo punto la frequenza, ma non con la stessa legge precedente. La taratura va effettuata a regime, e l'oscillatore deve rimanere sempre acceso dal momento della taratura, altrimenti perde il proprio senso. Si avrà sempre lo stesso coefficiente lineare α, ma un istante di assestamento diverso, e dunque una frequenza di partenza $f(t_0)$ diversa. Bisogna far dunque sì da avere anche una temperatura assolutamente stabile, al fine di non variare le caratteristiche meccaniche del quarzo (e di conseguenza, per piezoelettricità, quelle elettriche e di frequenza).

Al fine di tarare gli oscillatori, esistono sostanzialmente due tipi di campioni:

- Campioni primari: basati su costanti fisiche assolute e ben note, quali la velocità della luce c, o la costante di Planck $\hbar$; l'accuratezza dei campioni basati su questi fenomeni fisici, deriva direttamente dall'accuratezza con cui sono state determinate le costanti sulle quali ci si basa (o sulla conoscenza dei fenomeni fisici ad esse legati).

- Campioni secondari: legati in maniera indiretta ad altri fenomeni fisici che vanno misurati indirettamente; il campione secondario è dunque un campione tarato per confronto con un campione primario. L'accuratezza è solitamente inferiore, in quanto le cause di indeterminazione derivano sia dal campione primario, che dal metodo di confronto utilizzato per la realizzazione della taratura del campione secondario.

Fattore spesso da dichiarare riguarda la purezza spettrale del campione, o dell'oscillatore utilizzato; poichè l'oscillatore dovrebbe essenzialmente essere una sinusoide, idealmente nel dominio delle frequenze si dovrebbe visualizzare dunque una delta di Dirac, $\delta(f - f_0)$, dove f_0 sarebbe la frequenza del campione. Poichè ciò è assolutamente impossibile a causa dei soli fenomeni di rumore presenti, si definisce il parametro purezza spettrale, come la distanza in dB tra il segnale ed una riga spettrale distante Δf dalla frequenza f_0, con ampiezza di banda 1 Hz. Quelle che si dichiarano dunque per caratterizzare la purezza spettrale, sono la $\mathcal{P}_{dB}$, e l'offset Δf.

9.3 Sintetizzatori di frequenza

Abbiamo visto che gli oscillatori al quarzo hanno una frequenza molto stabile (se termostatizzati, ovviamente), ma fissa: questo grosso handicap ci impedisce di usarli come generatori di segnali (che dovrebbero essere al contrario strumenti estremamente versatili anche sotto il punto di vista delle frequenze). I sintetizzatori di frequenze sono strumenti in grado di generare sinusoidi nel range di frequenze dall'acustico alle microonde, con un'accuratezza piuttosto elevata, simile a quella del quarzo: questo perchè ciò che sta alla base di questi strumenti, in effetti, è proprio il quarzo stesso. Gli utilizzi di sintetizzatori di questo genere sono molteplici: dallo studio di funzioni di trasferimento di circuiti di vario genere, a sorgenti molto accurate e stabili, quali quelle per lo studio di radar, di effetti doppler in genere, di scansioni automatiche di frequenza.

Esistono fondamentalmente due metodi di sintesi di frequenza:

- Sintesi diretta: partendo dalla sorgente di riferimento a frequenza fissa (ossia l'oscillatore al quarzo finora definito), mediante operazioni di moltiplicazione per fattori N noti e divisione per dividendi M noti, applicate sulle frequenze, si ottiene una frequenza di uscita come combinazioni delle varie frequenze;

- Sintesi indiretta: si utilizzano dei PLL, ossia degli anelli ad aggancio di fase, e divisori numerici N, ed M, noti.

9.3.1 Sintesi diretta

Partiamo con l'analisi del processo di sintesi diretta di sinusoidi, ossia basandoci su di un riferimento fisso, che andrà combinato con altri ottenuti con moltiplicazioni e divisioni partendo dal primo. Innanzitutto, quello che serve introdurre è un dispositivo in grado di moltiplicare o dividere la frequenza della sinusoide introdotta in esso.

Si introduce la sinusoide in un circuito fortemente non lineare, ergo distorcente, che produce in uscita una serie di distorsioni armoniche; passando il segnale in un filtro passa banda, il cui massimo è situato in prossimità dell'armonica Nf_0, dove f_0 era quella di ingresso, si riesce ad ottenere di fatto una moltiplicazione di frequenza. Il duale di questo dispositivo, ossia il divisore di frequenza, funziona in maniera diversa: dividere le frequenze è molto più semplice, si può fare ad esempio mediante contatori (realizzati mediante porte logiche di vario tipo, o con circuiti utilizzanti retroazioni positive selettive: selezionando da un mixer una certa frequenza, e controreazionandola, si potrà ottenere una frequenza a propria scelta. Il sommatore di frequenze sarà basato su di un mixer: introducendo due frequenze f_1 e f_2, e filtrando in uscita dal mixer la frequenza $|f_1 + f_2|$ mediante un filtro passa banda, è possibile sommare di fatto le frequenze di due sinusoidi.

Riutilizzando dunque tecniche tratte dai precedenti tipi di generatori, ci è possibile ottenere uno strumento molto versatile, a patto di sviluppare anche le nostre conoscenze teoriche. Queste conoscenze risalgono alla II guerra mondiale, quando Mario Boella ideò il metodo della sintesi diretta, al fine di effettuare intercettazioni radar su frequenze piuttosto complicate, strane. Si creano le cifre della sequenza, rappresentante la frequenza del segnale sinusoidale che si intende sintetizzare, mediante le decadi di Boella. Si creano cifre di peso dapprima significativo e poi, man mano che si va verso l'uscita, di peso sempre meno significativo, mediante manipolazioni delle varie frequenze, utilizzando le operazioni di somma, sottrazione, moltiplicazione e divisione finora ottenute mediante i dispositivi appena introdotti. Il poter dividere per 10 permette di spostare le cifre di una posizione verso destra, e così combinare la cifra ottenuta con le frequenze che andremo a sommare o sottrarre, andando avanti con i blocchi dello strumento (moltiplicatori e divisori messi in particolari configurazioni, a seconda del dispositivo che si intende realizzare).

Più il sintetizzatore è complicato, ossia più sono i blocchi che lo compongono, e più esso avrà una risoluzione elevata. Limite massimo è dettato dalla stabilità del quarzo, che fornisce la frequenza base per il circuito. Questi sintetizzatori permettono una risoluzione dunque molto elevata, ma il fatto che servisse un numero così elevato di componenti li rende molto ingombranti, costosi, e difficili da realizzare.

9.3.2 Sintesi indiretta

Un'alternativa ai sintetizzatori basati sul processo di sintesi diretta sono i sintetizzatori indiretti, basati sull'uso dei PLL, ossia sugli anelli ad aggancio di fase (Phase-Lock Loop). Il principio alla base di questi dispositivi è il seguente: se due segnali sono isofrequenziali, allora la loro differenza di fase, $\Delta\Phi$, è costante; se non sono isofrequenziali, ovviamente, lo sfasamento tra i due segnali sarà variabile nel tempo, e quindi la differenza di fase non costante. L'idea interessante è sfruttare il segnale (variabile nel tempo) di differenza di fase $\Delta\Phi$, trasformandolo in un segnale di tensione (che avrà dunque una forma simile a quella di una rampa).

Il segnale di tensione così generato potrebbe comandare un VCO, ossia uno degli oscillatori comandati in tensione prima definiti, ottenendo il risultato che stiamo per presentare.

Utilizzando come riferimento un quarzo, e collegando il VCO ad un circuito comparatore di fase, passando per un amplificatore unito ad un filtro passa basso, si otterrà una tensione pseudocontinua. Il VCO continuerà ad inseguire la frequenza del quarzo, ossia la frequenza di riferimento, tendendo a stabilizzarsi con essa, per effetto della retroazione introdotta dall'anello comparatore di fase + integratore/passa basso + VCO. Dopo un certo tempo dunque, il VCO tenderà a stabilizzarsi (si noti che nelle modulazioni analogiche AM capita esattamente questo: si ha un oscillatore locale comandato in tensione che cerca di agganciare la frequenza della portante fino a raggiungerla, ottenendo quindi di fatto la portante pur senza bisogno di trasmetterla).

Complicando questo circuito, introducendo un divisore di frequenza dopo l'oscillatore al quarzo, ed un oscillatore di frequenza dopo il VCO, si riesce a sintetizzare frequenze diverse da quella introdotta precedentemente: cercando di spiegare in maniera più semplice, introdurre un divisore di frequenza pari a N in cascata al quarzo significa ridurre la frequenza del quarzo, e dunque diminuire la sua velocità, relativamente al PLL; introducendo a questo punto un divisore di frequenza pari a M in cascata al VCO, si tende a riavvicinare la frequenza a quella in ingresso al PLL, precedentemente normalizzata per N: in questo modo, data la frequenza di riferimento del quarzo non diviso, f_{rif}, si avrà che la frequenza in uscita al circuito, f_u, sarà pari a:

$$f_u = \frac{M}{N} f_{rif}$$

Quanto possiamo variare queste frequenze, utilizzando i divisori M e N ? Dipende dal PLL: il circuito passa basso dovrà lavorare a frequenza più bassa, l'anello aggancerà a frequenza più bassa la frequenza della sinusoide in uscita dal divisore, e quindi le fluttuazioni diventeranno più lunghe, e la qualità del circuito peggiorerà. Sarà necessario aumentare le costanti di tempo del circuito, aumentando la sua inerzia, in modo da permettere al PLL di agganciare, seppur lentamente, la frequenza in ingresso.

Questo tipo di sintesi è molto più economico da realizzare, mediante circuiti integrati (i PLL si trovano infatti integrati, anche assieme a circuiti divisori come quelli che utilizziamo), ed eventualmente un oscillatore al quarzo; l'introduzione di questi sistemi tuttavia riduce notevolmente la risoluzione rispetto al metodo di sintesi diretta, e anche la stabilità della frequenza che viene sintetizzata; per questo si utilizzano di solito per applicazioni economiche, quali selettori di canali per la televisione o per le radio, o generici usi di laboratorio (in misure dove non sono richieste elevate accuratezze).

Capitolo 10

Impedenzimetro Vettoriale

Oltre ai metodi basati su ponti, o sullo studio di circuiti risonanti, vi è un terzo sistema, un terzo strumento di misura di impedenze, basato sul metodo voltamperometrico: l'impedenzimetro vettoriale. Questo metodo consisterà sostanzialmente nel misurare, ad una certa pulsazione ω_0, i moduli della tensione V, della corrente I, e la loro differenza di fase Φ. Eccitata dunque l'impedenza mediante un segnale sinusoidale a frequenza ω_0, si misurano $|V|$, $|I|$, e Φ, ottenendo:

$$Z(\omega_0) = |Z(\omega_0)|\, e^{j\Phi} = R_0 + jX_0$$

Si tenga conto che è del tutto indifferente misurare i valori efficaci o massimi di V e I, ai nostri fini.

Per quanto riguarda le misure dei moduli delle correnti, si tratta di tecniche già utilizzate molto spesso; quello sul quale merita soffermarsi, è un'analisi delle misure di fase.

10.1 Fasometro numerico

Quella di cui abbiamo bisogno, al fine di realizzare uno strumento di misura di impedenze mediante misure di tipo voltamperometrico, è un metodo di misura diretta di fase; spesso, le misure di fase si fanno mediante strumenti analogici, nella fatispecie oscilloscopi analogici, dove, visualizzando i due segnali, si può vedere e misurare a occhio facilmente la differenza delle fasi.

Mai abbiamo sinora parlato di fasometri numerici, ossia di strumenti in grado di fornire un valore numerico dello sfasamento. Questo anche perchè in realtà la fase si presta male alla rappresentazione numerica: vedere su di uno strumento analogico la variazione della differenza di fase è molto più intuitivo rispetto al vedere un insieme di dati in ingresso in uno strumento.

Un'idea, è quella di utilizzare uno schema basato su dei comparatori di soglia, fissi ad un valore noto e semplice da utilizzare, come lo 0.

L'idea fondamentalmente è la seguente: dati in ingresso i due segnali, quando il primo fa commutare il comparatore di soglia, parte un segnale di START, che dà luogo ad un segnale di gate, un impulso; il secondo, quando fa commutare il comparatore di soglia (impostato alla stessa soglia del precedente, ossia a 0), provoca lo stop dell'impulso; relazionando la lunghezza dell'impulso rispetto a tutto il periodo T, è possibile calcolare la differenza di fase tra i due segnali. Le ampiezze del periodo T e dell'impulso t_0 vengono determinate mediante un processo di conteggio: abbiamo infatti, alla base di tutto, il solito oscillatore ad alta frequenza, introducente una serie di impulsi ad ogni oscillazione. Un circuito contatore sarà in grado di contare gli impulsi contenuti in t_0 ed in T, permettendo di stabilire dunque le seguenti relazioni:

$$N_0 \longleftrightarrow t_0$$

$$N_T \longleftrightarrow T$$

$$\Phi = 2\pi \frac{t_0}{T} = 2\pi \frac{N_0}{N_T}$$

Quello che si verrà dunque a creare, è un segnale onda quadra, rappresentante lo sfasamento tra i due segnali; di questo, è possibile estrarre la continua V_{DC}, ossia la componente continua del segnale onda quadra. Questa si definisce come:

$$V_{DC} = E_M \frac{t_0}{T}$$

Dove E_M è il valore di picco degli impulsi alti dell'onda quadra. Come si può notare, facilmente, $V_{DC} \propto \Phi$: le due grandezze variano una proporzionalmente all'altra. Dato dunque un fattore di taratura k, in grado di quantificare con maggior precisione il rapporto di proporzionalità tra le due grandezze, possiamo dire che:

$$\Phi = kV_{DC}$$

Siamo riusciti a realizzare quantomeno lo schema di principio del fasometro numerico. In realtà, ci sono alcuni problemi in esso che andrebbero rivisti e corretti: se $\Phi \simeq 0$, ossia lo sfasamento dei due segnali è quasi nullo, la presenza di rumore può provocare commutazioni indesiderate, e quindi l'impulso di gate potrebbe arrivare a durare per il 100% del periodo circa, falsando del tutto la nostra misura: del semplice rumore potrebbe infatti modificare i segnali, allungandone uno e di fatto invertendo le situazioni, ottenendo che il segnale in ritardo in realtà secondo il fasometro è in anticipo.

La soluzione a questo tipo di problemi è l'introduzione di uno sfasatore tarato: introducendo su uno dei canali uno sfasamento di 180 rispetto a quella che dovrebbe essere la sua fase, e quindi di fatto ribaltando il segnale, mettiamo i due segnali, che nella realtà sarebbero quasi coincidenti (in fatto di fase), in opposizione di fase, ottenendo un duty cycle del 50% circa. In un certo senso, dunque, abbiamo spostato l'indice a metà scala, eliminando il rischio di inversione del grafico.

Quello dei segnali simili tra loro non è l'unico problema che possiamo ritrovarci: se la frequenza dei segnali è elevata, infatti, a parità di sfasamento, il duty cycle potrebbe risultare più piccolo, in quanto i tempi di commutazione del circuito diventano non trascurabili. Inizia inoltre a farsi sentire il tempo di quantizzazione, ossia il limite imposto dalla frequenza dell'oscillatore di riferimento, dal momento che, se i segnali sono a frequenze elevate, il periodo sarà breve, il numero di impulsi contati piccolo, e si avrà una risoluzione inferiore. Ancora una volta ci verrà incontro lo sfasatore, questa volta tarato a scatti: a seconda dell'impostazione che introdurremo, potremo portare a centro scala il valore che noi desideriamo, evitando di incontrare anche problemi di questo tipo. Solitamente, come possiamo intuire, i fasometri numerici sono strumenti a zero centrale: la fase introdotta dallo sfasatore, dunque, deve essere in grado di portare i segnali in opposizione di fase a centro scala, in modo da ottenere una taratura del tipo:

$$\Phi_{tar} = 180 + \Phi_{CS}$$

In uscita a tutto, si introduce ancora un attenuatore tarato, in grado di riportare sulla scala che l'operatore esaminerà un valore accettabile.

10.2 Conversione di frequenza

Abbiamo un grosso problema, del quale non ci siamo ancora accorti: il circuito finora analizzato, funziona correttamente ma solo ad una determinata frequenza di lavoro, ad una certa frequenza di taratura. Qualsiasi sia la frequenza dei segnali in ingresso, il fasometro è in grado di lavorare esclusivamente ad una determinata frequenza ω_b.

Ciò che sarà necessario fare, è una conversione di frequenza, che mantenga inalterata la relazione di fase tra i segnali che analizziamo. Quello che si potrebbe fare, è il solito trucco del mixer: introducendo i due segnali a frequenza ω in due mixer, facendo in modo che le combinazioni lineari vengano effettuate con un generatore di frequenza variabile a frequenza ω_0, mediante due filtri passa banda si può fare in modo di selezionare, di tutte le le frequenze ottenute dalle combinazioni, quella pari a ω_b, ossia alla frequenza di lavoro del fasometro.

Mediante le formule di prostaferesi, si può verificare che di fatto, utilizzando un'operazione di questo tipo, si può mantenere la relazione di fase pur modificando le frequenze delle sinusoidi in ingresso.

Lo svantaggio di questo metodo è il fatto che si lavora sempre con un PLL, strumento adatto per lavorare ad alte frequenze; le variazioni che dovrà affrontare il dispositivo saranno molto elevate, e dunque i tempi di latenza da esso introdotti molto lunghi.

10.2.1 Tecnica di campionamento

Esiste un metodo alternativo a quello appena descritto, per ottenere lo stesso risultato: effettuare un processo di campionamento dei segnali. Nella fatispecie, campioniamo con gli stessi impulsi gli stessi istanti, ricostruendo quindi di fatto a bassa frequenza (utilizzando un sottocampionamento) il segnale.

Scegliendo dunque una frequenza di campionamento bassa, il segnale sarà ricostruito, a frequenza inferiore, mantenendo costante il campione. Utilizzando filtraggi, si riesce ad ottenere una curva continua, e liscia.

Sarà necessario scegliere un numero di campionamenti m, tale per cui:

$$T = m \cdot \tau$$

Dove τ è il ritardo tra le due sinusoidi. Più τ è corto, più il ritardo tra le due sarà piccolo, e quindi le curve meno sfasate tra loro, e più alta sarà la frequenza equivalente da utilizzare al fine di campionare in maniera corretta le curve.

Il periodo di campionamento, T_c, vale:

$$T_c = T_s + \tau$$

$$T_{eq} = mT_c = m(T_s + \tau)$$

Quindi:

$$T_s = m\tau \longrightarrow T_{eq} = T_s(m + 1)$$

La frequenza equivalente del segnale ricostruito, dunque, varrà:

$$f_{eq} = \frac{1}{T_{eq}} = \frac{1}{m+1} f_s = \frac{1}{\frac{1}{\tau f_s} + 1} = \frac{\tau f_s}{1 + \tau f_s} f_s$$

Alla base del fasometro, vi è un VCO, con però una particolarità: in questo ambito, la sua variazione è molto piccola rispetto a quella che di solito deve effettuare un oscillatore locale: sarà dunque molto più facile agganciare, in questo ambito, la frequenza. In uscita dal filtro avremo infatti da analizzare f_{eq}, normalmente intorno ad una qualche decina di kHz (frequenza di lavoro ottimale per il nostro fasometro).

Il fatto di sottocampionare, ci permette di effettuare misure di fase per qualsiasi tipo di frequenza: dovendo studiare sfasamenti in ambito di segnali a radiofrequenza, o sui GHz, il fatto che in ingresso al fasometro vi sia l'amplificatore selettivo, sottocampionante il segnale, non considera la frequenza di ingresso (a meno di eventuali frequenze di taglio provocanti distorsioni all'interno dell'amplificatore). L'utilizzo di questa tecnica, ha permesso di abbattere la barriera che rendeva impossibile la misura di segnali a frequenze dei MHz, rendendo tranquillamente realizzabili misure anche parlando dei GHz.

Possibili fonti di incertezza

La strumentazione finora presentata non è il massimo, sotto il punto di vista dell'incertezza: esistono diverse fonti di indeterminazione, che pregiudicano l'accuratezza delle nostre misure. Innanzitutto, a monte del fasometro, potrebbero esservi differenze di fase, ossia errori sul punto di zero, sull'azzeramento dello strumento. Questo errore è correggibile in maniera molto semplice: inviando lo stesso segnale ad entrambi gli ingressi dello strumento, ed impostando così lo zero a centro schermo, sarà possibile eliminare questo tipo di incertezza. Purtroppo questo è l'errore più banale e più semplice da risolvere: il fatto che si utilizzino attenuatori tarati, implica la possibilità di errori dovuti alla taratura; dal momento che inoltre si hanno due circuiti di condizionamento, con due differenti comparatori di soglia, è possibile che vi siano differenze tra le soglie definite nei due circuiti. Altro circuito che potrebbe presentare incertezza, è il formatore di impulsi, durante i quali si devono contare gli eventi generati dall'oscillatore. Si ricordi che bisogna saper prevedere e stimare le indeterminazioni dovute a tutti i contributi finora elencati, al fine di ben determinare il processo di misura.

10.3 Impedenzimetri vettoriali

L'introduzione al fasometro è fondamentale in quanto esso è l'elemento aggiunto, in grado di permettere ad un voltmetro di misurare la fase, oltre che il modulo.

Come già detto, alla base dello strumento vi è la tecnica voltamperometrica: si alimenta l'impedenza Z_x con un generatore di corrente sinusoidale a frequenza variabile; si divide a questo punto la misura in due fasi: misura del modulo, e misura della fase. Mediante un voltmetro in AC, si potrà misurare facilmente il modulo dell'impedenza, ossia la sua parte reale, ossia la parte puramente resistiva. Il fasometro misurerà la differenza di fase tra la tensione V sul carico e la corrente I circolante sull'impedenza. La tensione verrà dunque inviata esclusivamente al fasometro, mentre la tensione sia al voltmetro che al fasometro.

Sotto il punto di vista delle prestazioni, non sono molto basse: si ha la possibilità di misurare impedenze, su di un range di frequenze variabile da qualche decina di kHz a qualche centinaio di MHz (o GHz); si possono solitamente misurare impedenze da 0,1 Ω a 100 $k\Omega$, con un'accuratezza dell'ordine di $1\% \div 5\%$, come modulo, e $0,1 \div 2$. I dati di accuratezza introdotti sono però ottimali: se la taratura stata effettuata in modo da eliminare gli errori sistematici, allora possiamo sperare di aver effettivamente a che fare con incertezze dell'ordine di grandezza appena introdotto.

Con l'impedenzimetro vettoriale, è semplice dunque ottenere informazioni anche molto dettagliate riguardo la risposta in frequenza di circuiti RLC, e quindi caratterizzarli, eventualmente semplificarli, e fornire in prima approssimazione un andamento della loro funzione di trasferimento. Mediante l'impedenzimetro, infatti, è molto semplice valutare la pulsazione di risonanza del circuito, ω_0, la banda a -3 dB a partire da essa, e dunque le caratteristiche del circuito, utilizzandolo come una sorta di Q-metro.

Effettuando misure dettagliate, utilizzando come appena detto come Q-metro l'impedenzimetro, si tenga sempre conto dell'impedenza interna del generatore: la misura che si ottiene, è comprensiva anche degli effetti causati da essa, dunque se ne tenga conto, poichè la parte reale potrebbe causare abbassamenti del Q, e la parte reattiva una variazione dei valori di L o C del circuito.

Capitolo 11

Analizzatore di Spettro

L'analisi spettrale rappresenta l'esatto duale dell'analisi nel dominio del tempo: analizzare un segnale nel dominio delle frequenze, infatti, significa semplicemente rappresentare il segnale, considerando per ciascun punto dell'asse delle frequenze il contributo che le sinusoidi a quella frequenza hanno nella composizione del segnale. Si parla per questo motivo di spettro di frequenza di un segnale, attribuendovi il significato di rappresentazione, nel dominio della frequenza, di un certo segnale. Lo strumento di misura in grado di effettuare la presentazione di un segnale nel dominio del tempo è l'analizzatore di spettro: esso è in grado dunque di presentare lo spettro in frequenza (o la funzione di densità spettrale di potenza) di un segnale, su di uno schermo.

Come mai ricorrere ad un'analisi di questo tipo? La risposta è semplice: un fenomeno, un segnale, per essere analizzato in maniera semplice ed accurata, può essere analizzato sotto diversi punti di vista: un punto di vista è certamente quello che si può osservare su di un classico oscilloscopio, ossia il dominio del tempo: per poter analizzare quindi, come suggerisce il nome, la variazione nel tempo di una quantità, l'oscilloscopio (e quindi l'analisi temporale) rappresentano di sicuro il metodo più semplice. Supponiamo però di voler misurare alcune caratteristiche di un segnale, come ad esempio una distorsione, un rumore sovrapposto al segnale. Mediante l'oscilloscopio, si può vedere che al variare del tempo sono presenti disturbi sul segnale, ma non è assolutamente possibile studiarli o quantificarli; mediante un analizzatore di spettro, invece, tutto ciò risulta essere un'operazione molto semplice, studiando i contributi delle armoniche diverse dalla principale. Si può dunque facilmente quantificare mediante gli analizzatori di spettro i contributi di armoniche secondarie, sull'intero segnale, misurando la distorsione che introducono.

I segnali si possono caratterizzare in diverse maniere; una caratterizzazione idonea a ciò che dobbiamo attualmente fare, riguarda la loro stazionarietà, ossia la proprietà di mantenere (o meno) le caratteristiche (in questo ambito soprattutto si parla di frequenza) al variare del tempo. Al fine di studiare segnali stazionari, ossia le cui caratteristiche si mantengono inalterate nel tempo, o non stazionari, come transitori o segnali di vario genere, servono due strumenti di tipo fondamentalmente diverso:

- Per segnali non stazionari, transitori, sarà necessario uno strumento in grado di reagire molto velocemente alle variazioni del segnale, in modo da poterne captare il maggior numero di informazioni possibili. Lo strumento dunque deve essere molto veloce, nella fatispecie lavorare in real-time (ossia essere in grado di ottenere immediatamente lo spettro del segnale percepito), e così misurare tutte le componenti spettrali in tempi infinitesimi;

- Per segnali stazionari, o quasi-stazionari, è possibile utilizzare strumenti di tipo più semplice, semplicemente effettuando una scansione di frequenza (analizzatori sweep-tuned), ossia considerando filtri a sintonia variabile (o qualcosa di concettualmente simile, come vedremo meglio): il fatto che il segnale è stazionario, infatti, ci permette di poter effettuare con calma le operazioni, e quindi usare metodi di conseguenza più lenti.

11.1 Analizzatori real-time

Al fine di analizzare spettralmente segnali transitori o comunque generalmente spettro-varianti, sarà necessario, come già detto, un dispositivo molto veloce. Esiste la possibilità di costruire analizzatori di spettro analogici in grado di lavorare molto rapidamente, in real-time, tuttavia la nascita degli oscilloscopi analogici ha surclassato completamente questo genere di dispositivi. Dati segnali relativamente lenti (parlare di frequenza è improprio in quanto parliamo di fenomeni transitori, possiamo utilizzare il termine lenti per indicare il fatto che non si è in grado di rilevare componenti spettrali a frequenze molto elevate per limiti di vario genere dei dispositivi), si acquisisce, come in un oscilloscopio digitale, un certo numero di campioni nel dominio del tempo, e si inseriscono in una memoria. Mediante un algoritmo DFT, ossia di trasformata di Fourier discreta, FFT (Fast Fourier Transform), si riesce a trasformare i segnali nel dominio del tempo passando in uscita lo spettro in frequenza di questi. Mediante un CRT raster sarà possibile, proprio come nell'oscilloscopio digitale, presentare lo spettro ricavato in uscita dalla FFT.

Possiamo considerare questo tipo di strumento come un fratello maggiore del DSO: di fatto le operazioni che compie sono del tutto identiche, con in aggiunta la FFT: volessimo analizzare lo schema a blocchi dell'analizzatore di spettro numerico, infatti, avremmo gli stessi blocchi, più un processore ulteriore dedicato ad operazioni extra (quali proprio la trasformata discreta veloce di Fourier).

Questo tipo di dispositivo può lavorare benissimo (con le stesse prestazioni) anche con segnali di tipo stazionario; avremo anzi il vantaggio di poter campionare a frequenze inferiori, e definire ottimamente il segnale acquisito, prima di farne la trasformata discreta di Fourier. Oltre allo spettro di ampiezza, inoltre, è possibile determinare anche lo spettro di fase del segnale. Le limitazioni sono rappresentate fondamentalmente dalla velocità del campionatore, che deve essere in grado di memorizzare un certo numero di segnali ad una certa velocità, al fine di poter studiare correttamente il transitorio. L'accuratezza delle misure dipenderà da questo, e dall'algoritmo utilizzato per effettuare le operazioni di trasformazione, e di filtraggio numerico.

11.1.1 Analizzatori analogici real-time

Gli analizzatori analogici real-time, permettono di effettuare un'analisi spettrale immediata di tutto il segnale.

Questo tipo di strumento si realizza mediante un insieme di filtri passa-banda a banda stretta: ciascun filtro non farà passare alcuna banda (idealmente parlando) tranne quella interessata, ed in questo modo per ciascuna frequenza di ciascun filtro si ricava l'ampiezza; sullo schermo, si vedrà, come risultato finale, l'insieme delle righe spettrali ricavate mediante questo filtro. In uscita dal filtro, vi sarà un rilevatore di inviluppo: poichè il segnale sarebbe a frequenza molto elevata, e l'unica cosa che ci interessa è l'ampiezza, un rilevatore di inviluppo sarà in grado di fornircela senza problemi: l'ampiezza effettiva del segnale in ingresso è infatti diversa da quella di quella del segnale che vogliamo studiare, poichè il segnale in ingresso potrebbe essere stato modulato mediante modulazioni AM o FM ad esempio. Utilizzando questo sistema di demodulazione incoerente, potremo risalire alla componente effettiva del segnale modulante, e riportarla sullo schermo.

Dal momento che di fatto lo stesso segnale viene mandato in parallelo a ciascun filtro, che ne mantiene esclusivamente la parte che lo interessa direttamente, si ha un'acquisizione molto rapida del segnale. Al fine di aumentare la risoluzione, sarà necessario utilizzare più filtri, a banda stretta: in questo modo si potranno campionare con miglior risoluzione le varie righe spettrali, distinguendo meglio ciascuno dei contributi nello spettro.

Si noti una cosa: è necessario trovare un buon compromesso per quanto riguarda la larghezza di banda del filtro: dalla Teoria dei Segnali, sappiamo che vale il principio di indeterminazione della trasformata di Fourier: un segnale a supporto molto limitato in frequenza, avrà una risposta molto lunga nel tempo. Per questo motivo, utilizzando un filtro a banda molto stretta, si avrà una risposta nel tempo estremamente lunga, e quindi il dispositivo tenderà a non essere più real-time.

11.2 Analizzatori di spettro analogici

11.2.1 Analizzatori di spettro con filtri a sweep di frequenza

Cosa potremmo fare per realizzare un analizzatore sweep, cioè a scansione di frequenza? Un'idea potrebbe essere la seguente: utilizzare un filtro passa banda, stretto, a sintonia variabile. Questo filtro potrebbe scorrere sullo spettro, sull'asse delle frequenze, e visualizzare in questa maniera sullo schermo CRT lo spettro del segnale. Si otterrebbe così sullo schermo una sorta di funzione di trasferimento $|H(j\omega)|$. Sembrerebbe di aver realizzato lo strumento, e invece no: la soluzione appena presentata è assolutamente irrealizzabile, sotto il punto di vista pratico. Il filtro a sintonia variabile dovrebbe essere molto stretto e quindi, come nel caso precedente (analizzatori analogici di spettro in real-time), con una risposta nel tempo molto elevata. Dal momento che vi sarebbe un segnale in grado di modificare continuamente la frequenza del filtro, e dal momento che dunque il filtro è molto poco reattivo, l'uscita potrebbe non essere mai a regime, e dunque essere molto imprevedibile. Sarebbe necessario utilizzare un filtro più largo, in modo da ridurre la risposta nel tempo, ma perdendo così in risoluzione della misura. Altro modo sarebbe rendere molto molto lenta la scansione della frequenza, ma in questo modo si avrebbero scansioni estremamente lunghe, e quindi inadatte.

11.2.2 Analizzatori di spettro sweep-tuned a conversione di frequenza

Se Maometto va alla montagna, allora la montagna va da Maometto. Abbiamo capito che realizzare un analizzatore di spettro con un filtro a sintonia variabile è assolutamente impossibile, e anche se fosse possibile sarebbe a bassissima risoluzione, e con limiti notevoli. La soluzione di principio, ci ha fatto capire cosa deve accadere in teoria, ma in pratica dovremo rivoltare l'idea: il detto all'inizio del paragrafo, può farci intuire che la soluzione che permetterà di realizzare l'analizzatore sweep, anzichè usare un filtro che si muove nel dominio delle frequenze, sarà basata sull'uso di un filtro fisso, che analizza il dominio delle frequenze che si muove dentro di esso. In termini più tecnici, faremo qualcosa di molto simile a ciò che si usa nei generatori di segnali: il filtro passa banda a frequenza f_{IF} è fisso, mentre tutto lo spettro del segnale, la cui frequenza f_s è compresa tra due valori, $f_{s,min}$ e $f_{s,max}$, scorre sull'asse delle frequenze ν, mediante una continua conversione di frequenza basata sull'uso di un oscillatore locale (LO) la cui frequenza f_{LO} viene variata con continuità. Si parla di conversione di frequenza di tipo eterodina, poichè coinvolge due frequenze diverse, al fine di ottenere un'unica frequenza: le frequenze f_s e f_{LO} vengono mescolate, mediante un mixer, e di tutte le frequenze verrà selezionata solo quella del nostro filtro passa banda f_{IF}; supponendo ad esempio che la frequenza dell'oscillatore locale f_{LO} sia maggiore di quella del segnale, vediamo che verrà estratta:

$$f_{IF} = f_{LO} - f_s$$

Dove ricordiamo che lo spettro ha frequenza fissa f_s, appartenente alla banda $[f_{s,min}; f_{s,max}]$, e f_{LO} è variabile tra due valori (range di possibile variazione dell'oscillatore locale): $f_{LO,min}$ e $f_{LO,max}$; nella fatispecie, per stabilire il range di variazione della frequenza f_{LO}, si stabiliscono due picchetti:

$$f_{LO,min} - f_{s,min} = f_{IF}$$

$$f_{LO,max} - f_{s,max} = f_{IF}$$

In questa maniera, si riescono ad includere tutte le componenti spettrali di f_s, rendendole misurabili dal filtro a frequenza fissa f_{IF}.

Lo schema a blocchi potrebbe ricordare quello di un oscilloscopio analogico: si ha infatti un generatore di tensioni a rampa, che da un lato viene mandato al sistema di deflessione orizzontale di uno schermo CRT, e dall'altro va a fare da variatore di frequenza per l'oscillatore locale: sappiamo che quest'ultimo, infatti, sia un dispositivo comandato in tensione; utilizzando la rampa, si riesce a far assumere un certo range di valori a questo strumento, permettendogli di presentare una certa porzione di spettro. Anzichè basi tempi, si parla di una sorta di generatore di basi frequenze: la tensione a rampa che viene inviata all'oscillatore locale, lo comanda; per questo motivo, più sarà ripida la rampa, e più la scansione dello spettro sarà rapida, dal momento che variazioni veloci di tensione in ingresso all'oscillatore equivalgono a variazioni veloci di frequenza da esso regolate. L'introduzione di un attenuatore tarato tra generatore di rampa e oscillatore locale permette di variare il frequency span, ossia il range di variazione delle frequenze: l'attenuatore è in grado di modificare l'ampiezza delle rampe sull'asse delle ordinate, e così il range di frequenze nel quale può variare il nostro oscillatore locale. In uscita dall'oscilloscopio uscirà la nostra f_{LO}, mentre dall'ingresso dell'analizzatore di spettro l'altro segnale che verrà introdotto nel mescolatore, opportunamente filtrato e attenuato al fine di ridurre il rumore e rendere il segnale utilizzabile dal mixer. In uscita si avrà il contributo in frequenza dello spettro al punto di scorrimento determinato da f_{LO}, che verrà filtrato dal filtro alla sola f_{IF}, e mediante un rilevatore di inviluppo che rileverà l'ampiezza ed un filtro passa basso verrà mandato ad un attenuatore tarato. Questo, tarato nella fatispecie in dB (al fine di presentare sullo schermo uno spettro in scala logaritmica), permetterà di modificare l'ampiezza del segnale, al fine di regolare anche la sensibilità verticale, prima di essere così introdotto sul canale verticale dello schermo CRT. Mediante un controllo aggiuntivo, si ha la possibilità di variare la sensibilità, controllando la selettività del filtro passabanda IF: la banda di quest'ultimo infatti si potrebbe modificare mediante un controllo esterno (modificante ad esempio il Q del filtro).

Utilizzando strumenti di questo tipo, basati sulla conversione di frequenza e selettività di essa, si riesce ad ottenere buone prestazioni, in termini di rapporto segnale/rumore: il passa banda infatti è in grado di tagliare tutte le componenti di rumore nello spettro al di fuori di esso, ottenendo il risultato appena enunciato. Altro vantaggio rispetto agli altri strumenti finora presentati riguarda l'ampia gamma di scansione: la tecnica di conversione di frequenza permette una variazione relativa di frequenza molto elevata, a patto di utilizzare f_{LO} e f_{IF} superiori rispetto al range di frequenze che si intende scansionare (proprio come parlando di generatori di segnale).

11.2.3 Problematiche degli analizzatori sweep-tuned a conversione di frequenza

Purtroppo non tutto è semplice come abbiamo finora descritto: lo strumento appena introdotto ha alcune problematiche, alle quali bisognerà fare caso mediante alcuni accorgimenti.

Misure a bassa frequenza

Se si ha uno spettro con estensione anche a frequenze molto basse, prossime alla continua, e quindi $f_{s,min} \simeq 0$, avviene un fenomeno piuttosto spiacevole: dal momento che la frequenza viene misurata mediante differenza delle due frequenze, come già detto, quella che si misurerà, al momento di scansionare i contributi delle frequenze minime, sarà:

$$f_{LO,min} - f_{s,min} \simeq f_{LO,min}$$

A questo punto, passerà al filtro passa banda il contributo non della frequenza $f_s = 0$, ma il contributo di $f_{LO,min}$:

$$f_{IF} = f_{LO,min}$$

Quella che verrà visualizzata a questo punto sullo schermo dell'oscilloscopio non sarà una componente prossima alla continua dunque, bensì la componente legata al minor valore assumibile dall'oscillatore locale LO: il mescolatore non è in grado di rilevare linee prossime alla continua, ed il filtro cattura una frequenza errata. Questa frequenza, viene detta indicatore di frequenza zero, o oscillatore locale passante.

Frequenze immagine

Potrebbe capitare che, oltre allo spettro utile del segnale, si formi una sorta di spettro ombra, o spettro immagine: è infatti possibile che siano contenute alcune componenti immagine (ad esempio di rumore, o generiche righe spettrali inutili rispetto allo spettro del segnale), f_{IM}, più elevate delle f_{LO}, in grado di far sì che:

$$f_{IM} - f_{LO} = f_{IF}$$

Ossia, frequenze in grado di far raggiungere comunque il livello del filtro passa banda f_{IF}, considerando questa volta lo spettro presentabile dall'oscillatore locale come inferiore. Il rimedio a questo tipo di problema esiste ed è molto semplice da realizzare: introducendo all'ingresso dell'analizzatore di spettro un filtro passa basso, si eliminano le componenti spettrali non rientranti nel segnale da analizzare, ed in questo modo si elimina la possibilità di considerare a spettri immagine come quello appena descritto. La frequenza del filtro f_{IF} sarà di solito di poco inferiore alla $f_{LO,min}$, ed entrambi molto superiori alla frequenza di lavoro dello strumento (al fine di rappresentare un'ampia gamma di frequenze). La frequenza di taglio del filtro passa basso in ingresso, f_t, andrà scelta in modo da essere superiore alla $f_{s,max}$ del segnale, ma inferiore alla f_{IF}; in questo modo, si elimina il problema delle frequenze immagine.

Risoluzione in frequenza

La risoluzione in frequenza, come già detto, è la capacità di distinguere due componenti spettrali vicine tra loro. Questa, dipende dalla larghezza del filtro, esprimibile in termini di larghezza di banda (a -3 dB o a -6 dB, a seconda delle specifiche) del filtro IF (dette anche RBW_3 e RBW_6). Nella fatispecie, la risoluzione dipende da due fattori, fondamentalmente: distanza in frequenza, e distanza in ampiezza delle due righe spettrali.

- Se due righe hanno la stessa ampiezza, sarà sufficiente, al fine di distinguerle, che vi sia una differenza di ampiezza pari a 3 dB (o 6 dB a seconda delle specifiche); in questo caso, l'analizzatore di spettro si accorgerà della differenza, e presenterà sullo schermo un'unica riga. La possibilità di distinguere le due curve dipende dall'ampiezza di banda del filtro IF utilizzata: se utilizzassimo un filtro stretto abbastanza da permettere di visualizzare il varco in ampiezza a 3 dB delle due curve, sarebbe possibile distinguerle (sempre se effettivamente esse lo abbiano, questo varco).

- Se due righe sono a livello diverso, per distinguere le due righe spettrali, bisogna studiare il rapporto della banda a 3 dB con la banda che si intende vedere della curva. Perchè si possano distinguere, vi deve essere una differenza nel rapporto tra la banda che si desidera visualizzare (ad esempio supponiamo 60 dB) è necessario un filtro tale per cui:

  $$\frac{B_{60dB}}{B_{3dB}} \leq 6$$

  Per cercare di capire meglio, supponiamo di avere due righe spettrali sovrapposte: una è molto stretta in spettro ed alta in ampiezza, l'altra molto più larga in spettro e bassa in ampiezza rispetto alla prima. Se si utilizza un filtro molto largo, non si riesce a distinguere le due, poichè la riga stretta sembra soltanto essere il massimo della curva più larga. Restringendo la RBW_3 del filtro (considerando ossia un filtro dichiarato nelle specifiche con larghezza di banda a - 3 dB), possiamo distinguere sempre meglio la differenza tra le due curve. Il criterio che abbiamo utilizzato qua, è il seguente: sapendo che le due ampiezze differiscono dell'ampiezza A in dB che noi intendiamo presentare sull'oscilloscopio (nel nostro esempio 60 dB), se esse sono distanti in frequenza almeno X volte la larghezza di banda del filtro a frequenza IF, al fine di ottenere una corretta visualizzazione, dovrà essere verificata la relazione (valida per la banda 3/6 dB significa o a 3 o a 6 a seconda delle specifiche dello strumento):

  $$\frac{A_{dB}}{B_{3/6dB}} = 2X$$

Risposta in frequenza del filtro IF

Come già detto in precedenza, la risposta nel tempo del filtro IF è una cosa da non trascurare, poichè potrebbe determinare effetti molto spiacevoli. Al fine di effettuare correttamente le misure, serve che il filtro sia a regime: introdurre infatti un segnale che scorre nel filtro, richiede una sua continua risposta, e dunque un continuo transitorio. L'unico metodo per effettuare misure in maniera corretta di fatto è utilizzare scansioni estremamente lente, in modo da mandare a regime nel tempo il filtro, e così poter rappresentare correttamente lo spettro sullo schermo. Bisogna fare in modo da rendere la risposta nel dominio del tempo del filtro simile allo spettro: per quanto sia possibile, dobbiamo cercare di scansionare lentamente, ma comunque utilizzare una banda più stretta possibile. Gli analizzatori di spettro moderni, utilizzando una banda di filtraggio molto stretta, addirittura, ovviano questo problema bloccando le velocità di scansione al di sopra di un certo valore, in modo da non rovinare la presentazione su schermo.

Supponiamo di introdurre nell'analizzatore di spettro una sinusoide vicina all'idealità: ci aspettiamo che questo presenti sullo schermo una risposta di tipo impulsiva, ossia una singola riga in tutto lo schermo. Supponiamo di scansionare lentamente e a banda stretta tra la frequenza $f_{LO,min}$ e $f_{LO,max}$. Non è detto di fatto che si abbia in uscita, a queste condizioni, una riga spettrale, poichè ciò che si vede dipende dalla curva di selettività del filtro passa banda, ossia da quanto è effettivamente stretta e vicina ad un impulso la forma della funzione di trasferimento del filtro IF. Quello che capita nella realtà è che, essendo non nulla la banda intorno alla frequenza centrale del filtro, si mantengono delle sorte di residui, anche per le f_{LO} tali da non avere risposta spettrale. Nella realtà, quindi, non si dovrebbe vedere sullo schermo una riga, bensì una sorta di campana.

A nostro vantaggio tuttavia gioca il frequency span: poichè il range di frequenze esplorate solitamente è molto ampio, è possibile che questa sorta di campana venga visualizzata molto schiacciata, al fine di poter permettere una visualizzazione di una porzione molto ampia di spettro. Per questo motivo, in sostanza, impostando un frequency span sufficientemente elevato, si riesce ad ottenere, come risultato finale, una singola riga.

Ciò che abbiamo finora detto è tutto vero, a patto da eseguire la scansione molto lentamente, creando una situazione in prima approssimazione statica:

$$\frac{\Delta\nu}{\Delta t} \sim 0$$

In questa maniera, si fa in modo che, al momento della presentazione, l'uscita del filtro sia a regime, e così si ottenga una presentazione corretta dell'ampiezza nel punto interessato. Si può tuttavia quantificare, introducendo un parametro, la qualità delle presentazioni a partire da alcuni dettagli:

- La velocità di scansione, ossia la variazione di frequenza nel tempo:

$$\frac{\Delta\nu}{\Delta t}$$

- La larghezza di banda del filtro IF (consideriamo banda a -6 dB): B_6

Si definisce un parametro k, esprimente la qualità della presentazione, come:

$$k = \frac{\frac{\Delta\nu}{\Delta t}}{B_6^2}$$

All'aumentare del parametro k, possono accadere sostanzialmente due cose (intuibili facilmente leggendo la sua definizione):

- Aumenta la rapidità di scansione, e quindi la variazione nel tempo delle frequenze passanti entro il filtro, rendendo così più difficile il raggiungimento di un'uscita a regime;

- Diminuisce la banda passante del filtro, rendendo sì migliore la qualità dell'immagine, ma aumentando notevolmente la risposta nel tempo, e quindi rendendo meno reattivo il circuito (e allungando il regime).

Affinchè si abbia una corretta rappresentazione sullo schermo, si vuole che $k \ll 1$; per far ciò, è necessario operare sui controlli situati sul pannello dell'analizzatore di spettro:

- Ampiezza del range di frequenze da esplorare (frequency span);

- Velocità di scansione (o durata della scansione);

- Larghezza di banda del filtro IF.

Gli strumenti più moderni sono configurati in modo da impedire, se k sale al di sopra di una soglia troppo elevata, di utilizzare combinazioni dei tre parametri prima descritti (quindi ad esempio impedisce di alzare troppo la velocità di scansione).

Stabilità dell'oscillatore locale

Abbiamo finora parlato di regolare l'ampiezza di banda del filtro IF, senza tener conto di una cosa: la frequenza dell'oscillatore locale LO potrebbe essere molto instabile. Non avrebbe senso dunque utilizzare una banda troppo ridotta per il filtro passa banda: le fluttuazioni potrebbero cambiare la frequenza filtrata, e così non solo si otterrebbe una costante di tempo del sistema molto elevata, ma anche un risultato di fatto poco significativo. Si avrebbero dunque fluttuazioni delle frequenze rilevate, nella curva di selettività del filtro IF, trasformando il rumore in uscita significativa secondo il filtro, che presenterà la frequenza sbagliata a causa dell'instabilità dell'oscillatore.

Problemi di rappresentazione

Stiamo continuamente dicendo di cercare di ridurre la velocità di scansione, in modo da presentare sullo schermo lentamente i segnali. Ciò potrebbe sembrare furbo, ma in realtà si trascura un fatto: la tecnologia CRT si basa sulla persistenza dell'immagine su di uno schermo mediante fosfori: utilizzando fosfori molto lenti, ad elevata persistenza, sarebbe possibile ovviare per qualche momento a questo tipo di problema, ma non è una soluzione definitiva. La soluzione definitiva è utilizzare, quando è possibile, l'analizzatore di spettro digitale, che non ha di questi problemi; quando sarà necessario, ci si dovrà accontentare degli schermi che si hanno a disposizione in ambito analogico.

11.2.4 Conversioni multiple di frequenza

Abbiamo finora considerato un caso di singola conversione di frequenza: al fine di avere risoluzioni elevate, dovevamo stringere la banda del filtro fino a poche decine di Hz, cosa che è realizzabile, ma non a tutte le frequenze. Un filtro passa banda può essere realizzato con ampiezza di banda molto stretta, ma solo a frequenze basse, poichè alzandoci di frequenza iniziano ad esserci problemi legati al fattore di qualità del filtro. Utilizzando il principio di conversione singola di frequenza, siamo limitati dal fatto che:

$$Q = \frac{f_{IF}}{B}$$

Dove Q è un parametro che può essere a dir tanto 10^4. Volendo ottenere per esempio un'ampiezza di banda di 100 Hz, f_{IF} dovrebbe essere a 1 MHz, e quindi la banda di funzionamento dell'analizzatore di spettro sarebbe molto al di sotto del megahertz (poichè la conversione di frequenza è basata sull'utilizzo di filtri a frequenza più elevata del range di variazione permesso dallo strumento), limite che in effetti si preferirebbe poter raggiungere e superare.

Conversione doppia di frequenza

Quello che si è pensato, è effettuare, anzichè una singola conversione di frequenza, una doppia conversione. La prima è basata su di un filtro $f_{IF,1}$ con selettività molto bassa, e quindi una curva anche molto larga (poichè, come vedremo, la selettività al primo step non ci riguarda). Il segnale f_s viene mandato al mescolatore, assieme ad una $f_{LO,1}$ molto elevata e quindi si ha in uscita una componente spettrale a frequenza molto bassa, che chiameremo per l'appunto come il filtro IF, $f_{IF,1}$.

Abbiamo detto che a frequenze basse ci è possibile realizzare filtri anche molto stretti, per esempio con un'ampiezza di banda dell'ordine degli Hz. Realizzeremo dunque un filtro $f_{IF,2}$, a banda molto stretta (e quindi molto selettivo), nel quale introdurremo il segnale in uscita da un secondo mixer, mescolante $f_{LO,2}$ (più elevata di $f_{IF,2}$, e la frequenza in uscita dal primo mescolatore, $f_{IF,1}$. In uscita dal secondo filtro IF avremo dunque un segnale a banda molto stretta, tale per cui:

$$f_{IF,2} = f_{IF,1} - f_{LO,2}$$

La frequenza $f_{IF,2}$ è così bassa da rientrare nell'ordine di grandezza delle frequenze del segnale f_s da analizzare in principio. Utilizzando dunque semplicemente due volte il processo di conversione di frequenza, abbiamo ottenuto la possibilità di filtrare un segnale a frequenza molto bassa, e la tecnologia ci permette di effettuare, a frequenze basse, filtraggi più selettivi.

Come si può intuire, il primo filtro ha solo lo scopo di spostare le componenti spettrali che ci interessano a frequenze basse, e dunque la sua selettività non ha influenza nell'ambito della risoluzione finale della componente spettrale visualizzata: l'unico filtro veramente importante, sotto il punto di vista della risoluzione, è quello con frequenza centrale $f_{IF,2}$.

11.2.5 Rappresentazione zero span

Esiste una rappresentazione molto particolare, detta zero span (ossia, in cui la variazione di frequenza dell'oscillatore locale è pari a zero). Ciò significa che, dunque, f_{LO} è una costante: di fatto, l'analizzatore di spettro, funziona come un ricevitore sintonizzato su di una singola frequenza, con larghezza di banda definita dalla selettività del filtro IF. In questo modo, semplicemente, il segnale viene presentato nel dominio del tempo, come in un oscilloscopio.

Quello che entra nell'analizzatore di spettro, di solito, è un segnale modulato mediante modulazioni AM o FM; ciò che l'analizzatore di spettro effettua, analizzando il segnale in zero span, è estrarre, a partire dal segnale modulato, il segnale modulante e presentarlo sullo schermo, nel dominio del tempo, mediante un sistema di ricezione incoerente (ricordiamo che, in uscita dal filtro IF, vi è un rilevatore di inviluppo, in grado di captare il segnale modulato in ingresso e determinare il segnale modulante, al fine di tracciarne la corretta ampiezza senza che essa sia influenzata dalla modulazione utilizzata). Se l'analizzatore di spettro è digitale, è possibile a questo punto utilizzare una funzionalità molto interessante: campionati i vari punti del segnale modulante nel dominio del tempo, è possibile effettuare un'analisi spettrale del segnale modulante (e non del modulato, come si poteva aver fatto finora!), mediante un semplice algoritmo FFT. Si può così ottenere, in forma numerica, lo spettro del segnale modulante estratto mediante il rilevatore di inviluppo, in modalità zero span.

Capitolo 12

Lo standard IEEE-488

La nascita di strumenti molto più complicati di quelli tradizionali, quali ad esempio gli strumenti digitali (come l'oscilloscopio numerico), fece sorgere un grosso problema: ogni azienda produceva gli strumenti, utilizzando diversi componenti che venivano interconnessi tra loro mediante metodi di vario genere; verso la fine degli anni 60, ci si rese conto che era sempre più necessario uno standard per quanto riguarda l'interfaccia tra i vari componenti all'interno di uno strumento elettronico. Nel 1975 la IEEE pubblicò una norma, definente uno standard, chiamata 'Digital Interface for Programmable Instrumentation' (IEEE-488). Dopo di esse vennero pubblicate altre norme, ricalcanti il modello di quest'ultima e molto simili ad essa, tranne che in alcuni aspetti; di fatto gli standard erano però diversi, e collegare un dispositivo realizzato per un tipo di interfaccia mediante una di un altro genere, provocava danni ai circuiti delle interfacce. Per questo motivo tutti coloro che utilizzarono (specie in Europa) lo standard IEC-625, una delle altre norme uscite in quelli anni, dovette convertirsi allo standard IEEE-488. Questo nel 1987 venne ridefinito, migliorando alcuni aspetti, e ottenendo così lo standard ancora oggi utilizzato, IEEE-488.2 .

12.1 Caratteristiche fondamentali dello standard

Gli obiettivi del progetto, dello standard appena definito, erano quelli di definire un sistema di interconnessione di dispositivi su distanze limitate, al fine di poter integrare strumenti di diverso tipo, di diversa provenienza, in un unico sistema (come per esempio un oscilloscopio digitale); tutto ciò deve essere fatto in maniera ottimale, e dunque devono essere minimizzate le restrizioni che il sistema, ossia l'insieme dei blocchi, dei dispositivi, impone sulle prestazioni dei singoli strumenti: all'interno del sistema, ogni strumento deve poter lavorare al massimo, e non essere limitato da problemi per esempio dovuti all'interfaccia. La velocità dell'interfaccia deve dunque essere piuttosto elevata, in modo da non fornire un collo di bottiglia ai singoli dispositivi, permettendo una certa coralità, nel funzionamento globale del sistema.

Un sistema di interfacciamento è caratterizzato da alcuni aspetti, di diverso tipo, che lo riguardano, ed influenzano le sue prestazioni:

- Specifiche funzionali: riguardano la descrizione logica delle funzioni che comandano le attività dell'interfaccia, ossia il modo in cui sono logicamente definite le istruzioni che permettono di utilizzare l'interfaccia.

- Specifiche elettriche: stabiliscono la temporizzazione, le soglie elettriche di tensione o corrente, il tipo di logica che si utilizza per le specifiche funzionali, ecc.

- Specifiche meccaniche: riguardano la forma, le dimensioni, il tipo di connettore utilizzato, la sua posizione, la struttura fisica del cavo, ecc.

- Specifiche operative: l'insieme delle funzioni intrinseche di un dispositivo: il sistema di interfaccia serve a mettere assieme e far comunicare un certo numero di dispositivi, ma le specifiche operative riguardano ciò che ciascun dispositivo sia in grado di fare, slegato dall'interfaccia.

I sistemi basati sullo standard IEEE-488 sono a BUS, ossia utilizzano un insieme di fili paralleli per poter comunicare; ciascun filo rappresenta una linea, ognuna con un determinato scopo. Ci sono in totale 24 fili di BUS, ossia 24 linee, con i seguenti scopi:

- 8 linee dato, ossia 8 linee che servono esclusivamente per la trasmissione di dati;

- 8 schermi di massa, ciascuno per una linea dato;

- 3 linee di handshake (il cui scopo verrà meglio definito in seguito);

- 5 linee di gestione dell'interfaccia (idem, anche l'utilità di queste sarà meglio definita in seguito).

La logica utilizzata è TTL Schottky compatibile ($V_{LOW} \leq 0.8V$, $V_{HIGH} > 2V$), open collector o tri-state: in questa maniera, alcune linee sono gestite con una connessione wired-or, e così si riesce a ottenere un minor consumo di corrente nello stato logico FALSO. I messaggi di vario genere vengono trasferite sul BUS parallelo a 8 bit, in modo byte seriale, asincrono, e controllato dalla linea handshake a tre linee prima descritta (le modalità verranno meglio affrontate parlando di handshake nel dettaglio); il codice utilizzato per le trasmissioni, è l'ASCII standard versione 7 bit: anzichè l'ASCII extended a 8 bit; un bit viene utilizzato come bit di parità (come vedremo tra breve). L'architettura della rete, realizzata con questo tipo di interfacce, può essere a stella (tutti i dispositivi collegati allo stesso dispositivo, che nella fatispecie potrebbe essere un controllore), o a festone (il primo dispositivo è il controllore, e poi gli altri sono tutti collegati in serie). La lunghezza dei cavi è limitata, poichè si ha uno sviluppo totale massimo di lunghezza pari a $2N$, dove N è il numero di dispositivi collegati al BUS. A causa del fan-out delle porte logiche, il numero massimo di dispositivi collegabili è pari a 15; la velocità di trasmissione è dell'ordine di 1 MB/s come limite fisico, ma in realtà è difficile superare i 500 kB/s, a causa dell'handshake e dei dispositivi più lenti che partecipano ad esso.

Sulle 8 linee dati, vengono scambiati messaggi tra i vari dispositivi; essendo un sistema parallelo, tutti i dispositivi possono leggere tutti i dati, tutti i messaggi; i messaggi potrebbero essere:

- Device dependent: legati al tipo di apparecchiatura posta sul bus; a seconda dell'apparecchiatura, collegata al bus, vi saranno diverse codifiche dei dati: si potrebbe usare una codifica binaria, piuttosto che una codifica ASCII, piuttosto che altre codifiche;

- Device independent: si tratta di codici di sistema, standardizzati, che ogni dispositivo collegato deve essere in grado di interpretare. Questi sono i codici come già detto prima codificati mediante USA-ASCII a 7 bit (+ 1 bit di parità). Per ciascuna combinazione dei 7 bit corrisponde una certa casella, un certo numero, che permette di identificare un certo simbolo o una certa funzione dell'interfaccia.

I messaggi device independent sono molto importanti, proprio per questo motivo: essi costituiscono un insieme di comandi di sistema, ossia di comandi in grado di far lavorare il sistema, l'insieme dei vari blocchi, in una qualche maniera.

12.1.1 Ruoli dei dispositivi

Ciascun dispositivo all'interno del sistema può assumere un determinato ruolo, a scelta tra 4 particolari, determinati dallo standard di interfacciamento tra i dispositivi. Lo standard IEEE-488 prevede nella fatispecie i seguenti ruoli:

- Ascoltatore: il ruolo di ascoltatore è assegnato a ciascun dispositivo che deve ascoltare, ricevere i dati (device dependent) trasmessi sul bus. Semplicemente si tratta dunque di un ruolo intestante i ricevitori delle informazioni. Poichè vi deve essere almeno un parlatore (come vedremo tra breve), e poichè come già detto per limiti del fan-out delle porte logiche si possono avere al più 15 dispositivi collegati, al più potremo avere 14 ascoltatori in contemporanea.

- Parlatore: il ruolo di parlatore è il duale di quello di ascoltatore: si tratta del dispositivo incaricato alla trasmissione di dati (device dependent) in un certo istante di funzionamento del sistema. Questo ruolo può essere assunto da un solo dispositivo per volta (altrimenti si creerebbe confusione tra i vari dati trasmessi), di conseguenza è possibile che più dispositivi ascoltino lo stesso parlatore, ma non possono esserci in contemporanea più parlatori.

- Controllore: è un ruolo molto delicato, all'interno dell'interfaccia: si tratta di quel particolare dispositivo in grado di assegnare, mediante procedure di indirizzamento, a sè stesso o ad altri dispositivi i ruoli di parlatore o ascoltatore. A seconda delle richieste di servizio che gli vengono inviate, il controllore stabilisce i ruoli, e gestisce il funzionamento della rete. Si noti che in realtà, in sistemi molto complessi, vi possono essere più parlatori, anche se solo uno può essere attivo in un certo istante; viene definito system controller il particolare controllore in grado di controllare tutti gli eventuali controllori di un sistema: se il sistema come già detto è complesso, potrebbe presentare diversi controllori, nella fatispecie ciascuno per ogni sottosistema. Il system controller gestisce nel complesso il compito di gestione della rete, i controllori normali gestiscono solo alcune parti di rete, e possono essere fermati o accesi dal system controller.

- Ozioso: si tratta di un ruolo nullo, ossia di un tipo di dispositivo cui non è assegnato nè il ruolo di ascoltatore nè il ruolo di parlatore. In questo modo, si risparmia energia, e si evita di rallentare le operazioni di trasferimento dei dati. L'ozioso è comunque in grado di ascoltare dati a partire dal controllore, in modo da attivarsi ed assumere uno dei ruoli dettati dal controllore.

Si noti che in un sistem IEEE-488 potrebbe anche non esserci un controllore: se l'assegnazione dei ruoli è sempre statica, ossia non si ha mai un cambio di ruoli, allora non sarà necessario un dispositivo in grado di assegnare ruoli, poichè essi saranno sempre definiti dalla nascita dei vari dispositivi (che potranno dunque essere talker-only o listener-only).

Un dispositivo può avere più funzioni, distinguibili all'interno del dispositivo mediante alcuni comandi. Ogni dispositivo dispone di un indirizzo primario, in grado di identificarlo all'interno della rete, e permettere l'interfaccia col bus e quindi con gli altri dispositivi; è possibile che un dispositivo in grado di effettuare più funzioni abbia anche una funzione di indirizzamento secondaria o estesa: inviando oltre alla richiesta al primario un comando al sistema di indirizzamento secondario, sarà possibile selezionare la funzione che il dispositivo dovrà avere in un certo istante, e così il controllore potrà stabilire anche la funzione del dispositivo, oltre alla sua attivazione.

12.1.2 Handshake

L'handshake è una tecnica, una procedura, necessaria al fine di avere la garanzia che un segnale, un dato, sia inviato quando tutti i destinatari sono pronti a riceverlo, e per avere la garanzia che il dato venga effettivamente ricevuto dagli ascoltatori.

A questo fine, si utilizzano 3 linee di handshake, DAV (Data Valid), NRFD (Not Ready For Data), NDAC (Not Data Accepted). Cerchiamo di capire come funzionano queste tre linee, e quale sia il loro scopo (aldilà dei nomi, piuttosto esplicativi).

I dispositivi hanno bisogno di un certo tempo per ricevere il messaggio, devono essere tutti pronti: fino a quando ciò non è verificato, la linea NRFD è allo stato logico VERO, e dunque non si può avere trasmissione di dati. Dal momento che siamo in condizione wired-or, basta che uno solo dei dispositivi non sia pronto per ricevere, al fine di attivare la linea NRFD. Una volta che tutti i destinatari sono pronti, la linea NRFD viene posta su FALSO; a questo punto, il parlatore invia il primo data byte sul bus, e viene attivata la linea DAV. Ciascun destinatario a questo punto deve ricevere il primo data byte inviato dal parlatore, e disattivare la linea NDAC, indicante il fatto che non sono ancora arrivati dati a partire dalla trasmissione. Quando anche l'ultimo dispositivo avrà negato NDAC, si potrà riattivare NRFD, e così si riattiverà il ciclo: appena tutti saranno pronti, si disattiverà NRFD, si aspetterà un certo tempo, si attiverà DAV, poi si disattiveranno le NDAC, e così via.

Si noti che queste procedure di handshake si applicano sia per segnali device dependent che per segnali device independent: per i casi device dependent, l'handshake avverrà solo tra parlatore e dispositivi interessati a ricevere dati (mentre gli altri saranno sostanzialmente oziosi, mantenendo a

OFF i driver open collector); nel caso di comandi di sistemi, tutti i device dovranno invece partecipare all'handshake.

12.1.3 Messaggi multilinea

Come si sarà finora capito, sulle 8 linee dato possono transitare comandi (ossia segnali device independent), oppure dati (segnali device dependent): i primi devono essere codificati in maniera standard, ed essere comprensibili per qualsiasi dispositivo interno al sistema; per quanto riguarda i secondi, la cosa importante è che parlatore e ascoltatori siano in grado di comprendersi, e quindi i segnali devono essere codificati in modo da essere comprensibili.

Perchè si possano distinguere questi due tipi di messaggi, ossia il fatto che sulle linee di dato scorrano dati o comandi, si utilizza una delle cinque linee di gestione dell'interfaccia, ATN (Attention): se ATN è negata, allora i messaggi sul bus sono dati; se ATN è asserita, invece, i messaggi sui bus sono comandi. Dal momento che i comandi devono essere inviati dal controllore del sistema, allora ATN sarà gestita direttamente dal controllore attivo: quando ATN è attiva, tutti i dispositivi della rete, anche quelli idle, dovranno partecipare all'handshake; se ATN è disattivata, sarà necessaria la presenza all'handshake esclusivamente dei listener.

12.2 Comandi dello standard IEEE-488

Lo standard IEEE-488 prevede diversi tipi di comandi, trasferiti sul bus a 8 bit:

- Comandi di indirizzamento (in grado quindi di indirizzare il bus verso diversi device);
- Comandi universali (comandi che devono essere ricevuti in tutti i dispositivi);
- Comandi indirizzati (coinvolgenti un singolo dispositivo);
- Comandi secondari (riguardanti gli indirizzi secondari precedentemente citati).

12.2.1 Indirizzamento

L'indirizzamento è quell'operazione, effettuata dal controllore, in grado di attivare un dispositivo o come ascoltatore, o come parlatore. Questo tipo di operazione si effettua mediante un microswitch, o mediante controlli digitali presenti sul pannello dei sistemi. Cosa si fa? Ciascun dispositivo nella rete ha un certo indirizzo, in modo che ognuna abbia un indirizzo diverso da tutte le altre; il programma di gestione del sistema conosce ognuno degli indirizzi sapendo a quale interfaccia è assegnato, e così riesce a sapere cosa deve chiamare in un certo momento.

Se il controllore intende indirizzare sul bus un particolare dispositivo, allora introduce innanzitutto il bus in modalità comando (attivando la ATN), e inviando sulle linee di dato un byte (in realtà solo 7 bit, dal momento che la codifica ASCII prevede 7 bit di comando + 1 inutilizzato): i primi due bit stabiliscono quale sia il ruolo del dispositivo, e gli ultimi 5 stabiliscono l'indirizzo del dispositivo cui si vuole assegnare un certo ruolo. Si avrebbero a disposizione cinque bit di indirizzo dunque (pari a 32 possibili assegnazioni, anche se in realtà al massimo si hanno 15 dispositivi collegabili ad un bus), e due bit; in realtà, quando i bit di indirizzo sono tutti 1, e quindi si ha $Xxx11111$, dove X è il bit di parità, xx i bit di comando, e tutti gli altri sono 1, avviene una cosa interessante: si hanno a disposizione due comandi, legati ai due bit xx, detti UNT (untalker) e UNL (unlistener):

$$UNT = X1011111$$

$$UNL = X0111111$$

UNT è in grado di disindirizzare il parlatore, mentre UNL è in grado di disindirizzare tutti gli ascoltatori; una volta disindirizzati, i device entrano in status idle, ossia di oziatori, e aspettano di essere riutilizzati in una successiva fase del processo. Questo vale solo se gli ultimi 5 bit valgono tutti 1: solitamente questi indicano l'indirizzo specifico del device cui viene inviato il comando dettato dai primi due.

Comandi universali

Quelli appena proposti sono solo due dei comandi multilinea universali, inviabili ai vari strumenti: esistono in realtà sei comandi di questo tipo, ossia comandi che vengono inviati a tutti gli strumenti, ed effettuati da quelli che sono in grado di farlo. Proponiamone un elenco più completo:

- UNT (Untalk): disindirizza il parlatore attivo; in questo modo, prima di assegnare un nuovo parlatore, si può evitare che ve ne sia un altro attivo nel bus; il parlatore disindirizzato va in stato di idle.

- UNL (Unlisten): disindirizza tutti i dispositivi indirizzati come ascoltatori, portandoli in stato di idle.
- LLO (Local Lock Out): Disabilita l'impostazione manuale dei comandi sul pannello frontale di tutti gli strumenti in grado di riconoscere questo comando.
- SPE (Serial Poll Enable): abilita gli strumenti ad una risposta sequenziale in un'operazione di interrograzione, che segue ad una richiesta di servizio mediante serial polling (vedremo meglio in seguito).
- SPD (Serial Poll Disable): disabilita gli strumenti ad una risposta a richieste di tipo polling seriale.
- PPU (Parallel Poll Unconfigure): disabilita gli strumenti a rispondere a interrogazioni di tipo polling parallelo.

Comandi indirizzati

I comandi appena proposti vengono inviati a tutti i dispositivi collegati al bus, e interpretati da tutti i dispositivi in grado di effettuare le operazioni da essi richieste. Esistono altre classi di comandi che ora vedremo, tra cui questa: comandi che vengono letti ed eventualmente interpretati solo da tutti i dispositivi precedentemente abilitati come ascoltatori; tra questi vi sono:

- GET (Group Execute Trigger): tutti gli strumenti indirizzati come ascoltatori devono iniziare un'attività, determinata dalla loro impostazione. Quest'attività potrebbe essere per esempio un processo di misurazione, e ciò che fanno e le modalità sono specializzate nel dispositivo stesso: ogni dispositivo fa la sua parte, al fine di effettuare il processo richiesto dal sistema.
- SDC (Selected Device Clear): tutti gli strumenti indirizzati come ascoltatori vengono riportati ad uno stato di default, ossia ad uno stato predefinito.
- GTL (Go To Local): è il duale di Local Lock Out: fa in modo che tutti i dispositivi abilitati per essere programmati via bus, si possano programmare anche mediante i comandi manuali del pannello frontale (ossia in modalità locale).
- PPC (Parallel Poll Configure): configura il dispositivo, precedentemente già abilitato come ascoltatore, per istruirlo sulla modalità di risposta a operazione di interrogazione tipo polling parallelo.

- TCT (Take Control): viene usato dal controllore attivo, per trasferire il ruolo di controllore ad un dispositivo precedentemente indirizzato come ascoltatore.

Comandi unilinea

Esistono alcuni particolari messaggi, detti unilinea, che coinvolgono le linee di gestione dell'interfaccia; ciascuna delle linee viene utilizzata per inviare un particolare messaggio, legato allo stato della linea, tra i dispositivi della rete. Poichè si hanno cinque linee di controllo della rete, si avranno cinque segnali di controllo, uno per linea (uno di essi è il già citato ATN, in grado di differenziare il modo di trasmissione dati dal modo di comando). Vediamo lo scopo delle cinque linee:

- IFC (Interface Clear): ogni comunicazione sul bus viene interrotta, tutti i parlatori vengono disindirizzati e le operazioni di polling seriale disabilitate. Questo comando è utilizzabile solo dal system controller, ed in pratica il suo obiettivo è quello di bloccare tutte le operazioni che stanno avvenendo all'interno del bus, riportando ogni dispositivo in una situazione nota, iniziale.
- REN (Remote Enable): ogni dispositivo che è in grado di riconoscere questo messaggio, viene abilitato alla programmazione remota, ossia alla programmazione tramite bus (e tramite altri dispositivi presenti all'interno del sistema).
- SRQ (Service Request): i dispositivi presenti sulla rete chiedono attenzione da parte del controllore.
- ATN (Attention): il controllore, quando deve attivare il bus in modo comandi, ossia quando deve inviare comandi, la attiva. Tutti i dispositivi con ATN vera (compresi quelli in idle) interpretano i messaggi multilinea presenti sul bus come comandi; terminata l'operazione, si rimette a 0 la linea ATN da parte del controllore e la parola torna al parlatore, che continua a inviare dati ai soli ascoltatori.
- EOI (End Or Identify): questa linea ha due utilizzi, a seconda dello stato della ATN:
  - Se ATN=0 allora EOI è gestita dal parlatore attivo, ed indica la fine di una sequenza di dati inviati;

– Se ATN=1 allora il controllore cerca di identificare (Identify) quale è il dispositivo che ha richiesto un servizio (mediante SRQ) in una procedura di polling parallelo.

Quella appena proposta è solo una delle due possibilità di effettuare una terminazione dei dati, ossia di comunicare agli ascoltatori il termine di una trasmissione dati; un altro modo, consiste nell'inviare due particolari caratteri ASCII: CR (Carriage Return) + LF (Line Feed). Dal momento che si parla di dati, tuttavia, non è detto che gli ascoltatori siano in grado di interpretare perfettamente la codifica, di conseguenza sarebbe preferibile usare la EOI (con ATN=0, ovviamente, trovandoci in situazione di trasferimento dati), evitando maleinterpretazioni di questo tipo.

12.2.2 Richieste di servizio

Come abbiamo già accennato, parlando del comando SRQ, che i vari device possono effettuare una richiesta di servizio. La linea SRQ è gestita in wired-or, dunque è sufficiente che un singolo device faccia richiesta al fine di attivarla. Ci possono essere diverse motivazioni per un device per fare una richiesta di servizio: fornire dati, annunciare anomalie, o cose del genere. Purtroppo ci manca un dettaglio notevole, ossia il fatto che la linea SRQ può solo dire al controllore che vi è una richiesta di servizio, ma senza sapere precisamente da quale device arriva. Ciò che si deve effettuare è una procedura in grado di individuare lo strumento che ha richiesto il servizio. Questa operazione si effettua mediante un tipo di interrogazione, detto polling, dei vari dispositivi. Lo standard IEEE-488 prevede due tipi di polling, ossia il polling seriale, ed il polling parallelo.

Polling seriale

Il polling seriale effettua una richiesta serialmente, sequenzialmente, ad ogni device, un indirizzo per volta. Ciascun dispositivo che effettua una SRQ è in grado di abilitare un bit, indicante se è avvenuta o meno una richiesta di servizio. Gli altri bit sono in grado eventualmente, a seconda del dispositivo, di indicare cose diverse: una cosa ad esempio potrebbe essere l'eventuale motivazione per la richiesta di servizio. La procedura di polling seriale dunque procede così: prima di tutto il controllore invia il comando universale multilinea SPE, abilitando il polling seriale; ciascuno dei byte di stato dei vari device mediante una scansione sequenziale viene letto ed interpretato, in modo da poter riconoscere quali dispositivi han effettuato la SRQ: ogni dispositivo, uno per volta, viene assegnato come parlatore, mentre il controllore come ricevitore; il parlatore trasmetterà il proprio status byte al ricevitore, che stabilirà, leggendo il settimo bit (contenente le informazioni sull'eventuale inoltramento di una richiesta SRQ), se è stato effettivamente l'attuale parlatore ad aver richiesto. Una volta rilevato il corretto dispositivo, viene iniziata una fase di comunicazione allo scopo di determinare il motivo della richiesta, ed attivare procedure di servizio della richiesta. Si termina dunque, mediante SPD, la procedura di polling seriale.

Polling parallelo

L'inconveniente del polling seriale è la lentezza: esiste dunque un'alternativa all'accesso sequenziale a ciascuno dei dispositivi indirizzati al bus, ossia il polling parallelo: in un singolo step, si interrogano contemporaneamente tutti i dispositivi, rilevando immediatamente quale sia (o quali siano). Considerando il fatto abbiamo 8 linee, ipotizzando che per ogni linea vi sia al più uno strumento; ciascuna linea può essere o meno attivata (ossia assume valori 1 o 0 rispettivamente se è stato chiesto o meno un servizio dal dispositivo sulla tal linea).

Andiamo un po' più nei dettagli: prima di tutto il controllore abilita i dispositivi all'interrogazione mediante parallel polling, inviando prima il comando PPC (che configura secondo polling parallelo i device), e poi PPE (abilitante effettivamente il parallel polling); in quest'ultimo comando, sono contenuti negli ultimi tre bit gli indirizzi delle varie linee di comando (da 0 a 7, dal momento che ci sono otto linee son sufficienti tre bit per esprimerle tutte), ed il sense della risposta (si introduce come quart'ultimo bit un 1 o uno 0, a seconda se vogliamo che il dispositivo richiedente servizio risponda rispettivamente con 1 o 0). Il controllore, una volta abilitata la possibilità di effettuare un polling, abilitando contemporanealmente la linea ATN e la linea EOI, effettuerà l'identificazione (rivedere la descrizione del comando EOI): questa varrà come interrogazione a tutti i dispositivi abilitati al polling parallelo, che resituiranno un bit, per un totale di un byte (8 bit) ricevuti. Interpretando questo byte, il controllore potrà identificare quale linea ha effettivamente avanzato la richiesta, e così individuare chi ha mandato la SRQ.

Una volta terminata l'interrogazione, si deve terminare la procedura di polling parallelo mandando il comando PPU, in grado di deconfigurare i device riguardo questo tipo di interrogazione.

Capitolo 13

Analizzatore di stati logici

Progettare circuiti elettronici digitali o sistemi hardware/software non è assolutamente semplice: molto spesso il progettista potrebbe commettere errori di vario genere, sia nella progettazione dell'hardware, che nell'integrazione tra hardware e software. Non solo: sistemi elettronici di questo genere sono soggetti ad una certa usura, per vari motivi; queste sono le cause che hanno portato alla necessità di ideare uno strumento elettronico in grado di analizzare lo stato digitale di un determinato sistema elettronico: l'analizzatore di stati logici.

Un sistema digitale, data l'essenzialità delle informazioni trasportate ed utilizzate per la determinazione di risultati (esclusivamente bit, ossia combinazioni di uni e zeri), utilizza molti fili per la trasmissione delle informazioni; il fatto che si abbia a che fare con informazioni transitorie, sia per quanto riguarda i dati regolari trasmessi che le anomalie, e il fatto che non si punti più a studiare pochi segnali complessi, quanto un grosso numero di segnali binari, rende inadatto l'uso di strumenti classici, quali quelli utilizzati per i circuiti analogici. Sarà dunque necessario un nuovo tipo di strumento, più adatto al genere di esigenze introdotte da circuiti di tipo digitale.

Parlando di logic analyzer, o analizzatore di stati logici, spostiamo dunque notevolmente le nostre richieste rispetto a quelle che avevamo con uno qualunque dei dispositivi finora analizzati: non ci interessa più effettuare misure accurate di tensione, o di corrente, dal momento che ci interessa esclusivamente valutare se un bit vale 0 o 1, interpretando se i livelli sono High o Low.

Potremmo pensare all'analizzatore di stati logici come una sorta di oscilloscopio digitale, concettualmente parlando, ma molto più specializzato: esso deve avere un numero molto elevato di ingressi, al fine di poter acquisire molti dati, relativi a molte differenti linee, per esempio nel caso di circuiti integrati, in modo da definire correttamente lo stato del circuito; deve essere in grado di interpretare un valore di tensione con risoluzione 1 bit: deve essere quindi in grado di distinguere semplicemente 1 da 0. Dati n di questi ingressi, deve dunque essere in grado di presentare in uscita una sequenza di n stati logici, o di bit, o mnemonici legati ad un certo stato di bit: dal momento che trattiamo di fatto circuiti digitali, potrebbe risultare utile collegare, secondo convenzioni, determinate sequenze di bit con dei nomi, in grado di permetterci di riconoscerle facilmente; questi nomi vengono comunemente detti mnemonici. Dal momento che dobbiamo analizzare stati, in diversi momenti, servirà una memoria molto grossa, in grado di immagazzinare un grosso numero di dati (e quindi di stati); servirà infine una certa versatilità, nella scelta delle configurazioni di trigger, in modo da definire in modo molto versatile finestre temporali di osservazione dei dati. Serve una buona versatilità anche sotto il punto di vista delle funzioni utilizzabili, dal momento che in sistemi digitali potrebbero presentarsi anomalie di diverso tipo, come extra-bit dovuti a transizioni di un ingresso durante lo stesso periodo di clock (glitch).

13.1 Struttura e funzioni

L'analizzatore di stati logici come già detto rappresenta una specializzazione di un oscilloscopio digitale: alla base di tutto vi sarà un sistema di sonde e comparatori, in grado di rilevare in diversi punti del circuito diversi stati, e codificarli in una codifica binaria (1 o 0); questi valori verranno campionati, ed inseriti in una memoria tipo FIFO (ossia un registro a scorrimento); essendo FIFO, una volta riempita la memoria, verranno esplusi i primi valori inseriti, ed il contenuto di memoria continuerà, di qui, ad aggiornarsi. Esistono sostanzialmente due modalità di funzionamento di un analizzatore di stati logici:

- Modalità serie: si acquisisce il segnale su di un'unica linea, e così si campionano i valori su di una stessa linea, acquisendo n campioni ed ottenendo in uscita una singola parola di n elementi, contenente gli n stati della linea studiata;

- Modalità parallelo: viene effettuata un'acquisizione molteplice su N linee, in modo da ottenere N parole rappresentanti gli stati di N linee del circuito (dando quindi, volendo, un'idea più ampia rispetto alla modalità serie di quello che sia lo stato del circuito); è molto più utilizzata, normalmente, della modalità serie.

Altra classificazione delle modalità di funzionamento di un analizzatore di stati logici viene effettuata in base al dominio di funzionamento dello strumento: esistono infatti, sotto questo punto di vista, due possibilità:

- Dominio del tempo (modalità timing);
- Dominio dei dati (modalità data).

13.1.1 Modalità timing

Nella modalità timing, i segnali acquisiti vengono campionati ad una frequenza fissa, generata all'interno dell'analizzatore mediante un qualche circuito di clock, modificabile dall'utente. Ciò che viene dunque presentato dall'oggetto è un segnale binario funzione del tempo, ossia una forma d'onda binaria al variare del tempo.

A seconda di quante linee si analizzino al contempo, verranno tutte presentate, su di assi separati, identificate da etichette o dal numero identificante il canale cui è collegata la sonda analizzante la linea.

Volendo dunque avere una visione di insieme dei segnali, e valutare le relazioni temporali tra questi, ossia poter studiare la variazione dell'insieme degli stati al dominio del tempo, sicuramente questo tipo di rappresentazione sarà la più idonea.

La risoluzione temporale Δt, ossia la minima variazione di stato presentabile, dipende dal periodo di campionamento T_C, o meglio addirittura arriva a coincidere con esso:

$$\Delta t = T_C$$

Ciò significa che la minima variazione di stato percepibile, col variare del tempo, coincide con la frequenza di campionamento.

Nel caso uno degli impulsi di glitch (di cui prima abbiam fatto cenno) durino per un tempo percepibile dal campionatore, sarà sicuramente possibile vederli mediante questo tipo di tecnica.

13.1.2 Modalità data

Quando si studiano circuiti digitali sincroni, ossia le cui variazioni sono regolate da un circuito di clock, è possibile effettuare una piccola astuzia: utilizzare per il segnale di campionamento, lo stesso clock del circuito sincrono, configurando l'analizzatore di stati in modo da campionare in prossimità degli eventi di commutazione del circuito (questo a seconda della topologia del circuito digitale in studio). Per esempio, avendo un circuito in grado di commutare solo durante i fronti di salita o discesa, sarà sufficiente impostare gli istanti di campionamento solo in prossimità dei fronti di salita o discesa del clock, ottenendo un analizzatore di stato sincrono al circuito, il cui comportamento è legato al clock dell'analizzatore stesso.

Ciò che si rileverà a questo punto non sarà più un segnale al variare del tempo, poichè il tempo del circuito e quello dell'analizzatore sarà lo stesso: avremo dunque perso qualsiasi riferimento temporale, e sullo schermo presenteremo semplicemente una sequenza di uni e zeri, rappresentante l'evoluzione degli stati. Al posto di presentare questi, come già accennato, è possibile presentare mnemonici, per esempio utilizzando una codifica tipo ASCII. Addirittura, conoscendo gli strumenti che si stanno studiando (quali i processori INTEL 8086), è possibile utilizzare l'analizzatore di stati logici come riconoscitore di istruzioni, analizzando ciò che il processore sta facendo, e presentando in output la relativa istruzione assembly.

Questo tipo di modalità di funzionamento si può utilizzare sia in seriale che in parallelo: in questo modo si potranno osservare una o più linee, e studiarne lo stato (a seconda delle proprie esigenze).

Il problema che si ha a questo punto è il seguente: dal momento che il segnale di clock è esterno all'analizzatore di stati logici, e spesso dunque non periodico, potrebbe far campionare un certo numero di elementi di fatto non interessanti; ciò che si può fare per evitare ciò è introdurre un segnale ulteriore, in grado di abilitare l'acquisizione solo per intervalli di tempo stabiliti dall'utente, mediante un segnale di qualificazione del clock, che permette di rendere valido il segnale di clock per il campionamento solo a condizioni determinate dall'operatore dello strumento.

13.2 Sonde

Abbiamo parlato di presentazione, e funzionamento di questo tipo di strumento; come si fa tuttavia a catturare i segnali che poi dovremo presentare? Molto semplice: utilizzando sonde, i cui terminali sono progettati in modo da attaccarsi ai pin del circuito che intendiamo studiare. Dal momento che capita di studiare un numero molto elevato di stati, spesso le sonde vengono fornite in scatole, dette pod, ingrado di raggrupparne qualche centinaio, collegandole con un cavo multipolare a ciascuno degli ingressi dell'analizzatore di stati logici.

In questi ambiti stiamo parlando di microelettronica, poichè andiamo a utilizzare sonde di questo tipo in ambito di circuiti integrati; le sonde tuttavia possono introdurre effetti molto indesiderati, quali perdite, che potrebbero compromettere la misura: tipicamente, infatti, le sonde hanno una resistenza dell'ordine di $0.1 \div 1$ MΩ, con in parallelo una capacità di $10 \div 20$ pF; ciò potrebbe fungere da filtro passa basso, tagliando parte delle armoniche e limitando la velocità dei segnali (specialmente nel caso di logiche veloci). Quando vogliamo utilizzare dunque per l'acquisizione di dati sistemi di campionamento a frequenza molto elevata, sarà necessario utilizzare sonde compensate (concettualmente simili a quelle che si utilizzavano nell'ambito degli oscilloscopi). I pod al loro interno contengono i comparatori di soglia, che effettueranno il riconoscimento di un segnale, rilevando se esso vale logicamente 0 o 1; dal momento che la soglia non è univoca, a seconda del tipo di famiglia logica che si utilizza (CMOS piuttosto che TTL piuttosto che ECL), spesso è introdotto nel pod un sistema in grado di selezionare la soglia del comparatore, in modo da adattarsi alle specifiche del sistema in analisi. Nel caso di linee di pod molto particolari, è addirittura possibile regolare con continuità la soglia, al fine di determinare informazioni anche per quanto riguarda sistemi molto particolari; altro motivo per cui variare con continuità può essere utile è controllare se vengono verificati overshot, ossia margini di soglia che in teoria non dovrebbero essere previsti al fine di far funzionare bene il sistema elettronico (il verificarsi di overshot potrebbe infatti deteriorare notevolmente il sistema elettronico, riducendone il tempo di vita).

13.3 Particolari circuiti di campionamento

Abbiamo tendenzialmente spiegato come funzionino i circuiti di campionamento interni allo strumento: dal momento che essi commutano sostanzialmente con i fronti del clock, possiamo intuire facilmente che si tratti di flip-flop (nella fatispecie tipo D), ossia dispositivi in grado di catturare un segnale sull'ingresso D quando il clock ad essi collegato ha un fronte in salita.

Ciò che limita questo tipo di campionamento è la presenza dei segnali di errore da noi tanto temuti, i glitch: essi hanno una durata che potrebbe essere inferiore a quella ottenibile mediante il circuito di campionamento finora descritto, di conseguenza sarà necessario utilizzare uno stratagemma di tipo differente, cambiando l'elemento critico del circuito di campionamento: il flip-flop. Utilizzando un circuito digitale basato su dei latch, ossia circuiti bistabili in grado di commutare sulle transizioni, mantenendo costante lo stato in uscita per un tempo superiore al tempo di campionamento, sarà possibile allungare la visibilità dell'evento di glitch, campionarlo, e così rilevarlo e poterne effettivamente determinare la presenza. Questo tipo di stratagemma, basato sull'uso di latch, è utilizzabile ogni qualvolta si abbia il sospetto di avere su di una linea impulsi non visualizzabili a causa dei limiti della logica a flip-flop: confrontando i dati ottenuti con circuito a latch e circuito a flip-flop si riesce a comprendere se il dubbio era fondato o meno.

13.4 Sincronizzazione

Come nel caso dell'oscilloscopio digitale, di cui come già detto più volte questo non rappresenta altro che un caso particolare, un'applicazione di tipo più specializzato al fine del semplice studio dei sistemi elettronici digitali, c'è la necessità di posizionare la finestra di osservazione (ed i dati immagazzinati in memoria) nell'intorno di un istante temporale stabilito dall'operatore della misura, in modo più flessibile. Esiste come nell'oscilloscopio un sistema di trigger, in grado di condizionare, a partire da un particolare evento, l'arresto dell'acquisizione ed il congelamento in memoria dei dati che devono essere presentati sulla periferica di output. Dal momento che quello che stiamo realizzando è un analizzatore di stati logici, quello che si può fare è congelare gli stati una volta rilevata una condizione particolare proprio su di essi: dato il riconoscimento di una particolare sequenza, o di una particolare combinazione di variabili/stati, è dunque possibile inviare un segnale di trigger in grado di bloccare l'acquisizione e permettere all'operatore di studiare l'evento interessato.

Poichè uno dei motivi sostanziali per cui è stato realizzato questo tipo di strumento è il rilevamento di errori, sfruttando i metodi prima proposti di campionamento con ciò che è stato appena detto per quanto riguarda il trigger, è possibile combinare le due cose e così legare eventi di trigger ad eventi di errore, in modo da rilevare e studiare malfunzionamenti di vario genere.

Si ha ovviamente una flessibilità notevole per quanto riguarda l'arresto dell'acquisizione e immagazzinamento di dati: è possibile fare in modo da arrestare l'acquisizione prima, dopo nell'istante dell'evento di trigger, e presentare sullo schermo ciò che è stato registrato a discrezione dell'utente.