

Misure a radiofrequenza

Alberto Tibaldi

12 dicembre 2010

Indice

1	Richiami di Campi Elettromagnetici	4
1.1	Concetti fondamentali sulle guide d'onda	5
1.1.1	Guida d'onda rettangolare	6
1.1.2	Cenni alle discontinuità in guida	10
1.1.3	Cavo coassiale	11
2	Generatori a microonde	14
2.1	Concetti fondamentali per generare segnali a microonde	14
2.1.1	Purezza spettrale: phase noise	15
2.1.2	Altre caratteristiche	16
2.1.3	Circuiti risonanti	17
2.2	Realizzazioni di oscillatori	18
2.2.1	Oscillatore basato su cavità risonante	18
2.2.2	Varactor - richiamo al calcolo della sensibilità	19
2.2.3	YIG: Yittrium Iron Garnet	20
2.2.4	DRO: Dielectric Resonators Oscillators	22
2.2.5	Stadio di amplificazione	24
2.2.6	Oscillatori a diodo Gunn	28
2.2.7	Sintetizzatori a microonde	28
2.2.8	Divisione per sottocampionamento	30
2.3	DDS: Direct Digital Synthesis	36
2.3.1	Soluzioni integranti DDS e PLL	40
2.3.2	Stabilizzazione di ampiezza	41
2.3.3	Catalogazione storica dei sintetizzatori	42
2.4	Analisi di schemi di generatori	43
2.4.1	Analisi dello schema del generatore 87MX	43
2.4.2	Analisi dello schema del generatore Anritsu MG369XA MM	44
2.4.3	Analisi dello schema del generatore Agilent E8257N	49

3	Analizzatori di rete	51
3.1	Introduzione agli analizzatori di rete	51
3.1.1	Schemi di base dei VNA (Vector Network Analyzers)	58
3.2	Introduzione agli schemi fondamentali	60
3.2.1	Schema Anritsu 12 porte 65 GHz	61
3.2.2	Schema Agilent	61
3.2.3	Ipotetico schema	62
3.3	Analisi di alcuni schemi di network analyzers	63
3.3.1	Analisi dello schema del PNA	63
3.3.2	Analisi dello schema del 8720	67
3.4	Taratura degli analizzatori di rete	70
3.4.1	Problematiche introduttive	70
3.4.2	Introduzione alla taratura	71
3.4.3	Standard di calibrazione	72
3.4.4	Definizione del modello di errore	76
3.4.5	Modello di errore a due porte	83
3.4.6	Principali tecniche di taratura	84
3.5	Tecniche di taratura	86
3.5.1	Approccio basato sulle matrici di trasmissione	86
3.5.2	SOLR: Short, Open, Load, Reciprocal	98
3.5.3	Approccio basato sul modello di errore forward-reverse	99
3.5.4	SOLT applicata a un multiporte	101
3.6	Taratura di analizzatori di reti multiporte	102
3.6.1	Formulazione relativa ad analizzatori 4-sampler	103
3.6.2	Formulazione relativa ad analizzatori 3-sampler	106
3.6.3	Calcolo degli error coefficients	109
3.6.4	Cause di incertezza	112
3.6.5	Tecniche avanzate di taratura di multiporte	119
3.6.6	E-CAL: Electronic Calibration	122
4	Introduzione ad alcuni strumenti di misura	123
4.1	Analizzatore di spettro	123
4.1.1	Analisi del FSU Rohde & Schwarz	125
4.2	Power meters	129
4.2.1	Introduzione	129
4.2.2	Power meter a microcalorimetro	130
4.2.3	Power meter a termocoppia	131
4.2.4	Power meter a diodi	132
4.2.5	Accuracy dei power meter	133
4.3	Noise Figure meter	134
4.3.1	Richiami sul rumore	134

4.3.2	Hot-cold method	136
4.3.3	Strumento per la misura del rumore: il noise figure meter	138
5	Caratterizzazione di dispositivi di potenza	139
5.1	Introduzione al load pull	140
5.2	Load pull passivo	141
5.2.1	Real-time load pull	142
5.2.2	Uso di wave probe	145
5.3	Load pull attivo	146
5.3.1	Metodo Takayama per il load pull	150
5.3.2	Fixtures	152
5.4	Time-domain analysis: Non-linear Network Analyzer	153

Capitolo 1

Richiami di Campi Elettromagnetici

In questo capitolo si intende riprendere un certo numero di nozioni riguardanti la materia “Campi Elettromagnetici”, nozioni molto importanti al fine di poter discutere a un certo livello di strumentazione a radiofrequenza.

Si sappia da subito una cosa: quando si parla di misure a radiofrequenza, si pensi all’andare dal salumiere, e dato un piatto carico con una tonnellata di qualcosa, metterci sopra ancora 100 grammi di prosciutto; detto così sembra assurdo ma si supponga di aver a che fare con la realizzazione di uno strumento in grado di avere le seguenti performance sotto il punto di vista degli errori sistematici: $\lambda = 2$ cm, $l_{\text{cavo}} = 1$ m, precisione di misura della lunghezza elettrica di 1° . È complicato fare qualcosa di questo genere? Facciamo due conti:

- in 1 m di cavo vi sono, data $\lambda = 0,02$ m, 50 lunghezze elettriche:

$$\frac{l}{\lambda} = 50$$

- in ogni lunghezza elettrica vi sono 360° ;

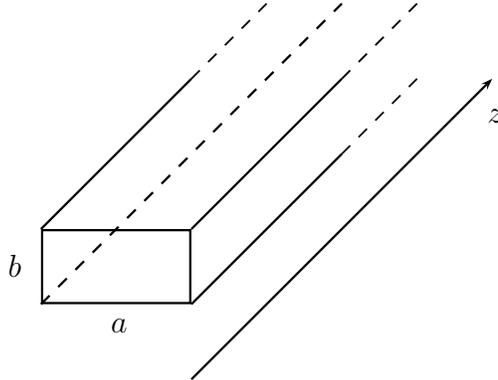
facendo un banale conto,

$$50 \times 360^\circ = 18000^\circ$$

e noi dobbiamo misurare con la precisione di 1° : questo problema è piuttosto simile, in termini di rapporti, a quello del povero salumiere: apprezzare 1° su 18000° non è assolutamente banale da fare, e per farlo bisognerà prestare molta attenzione a molti dettagli. Torneremo a ciò in seguito.

1.1 Concetti fondamentali sulle guide d'onda

L'approccio più usato nell'ambito dell'ingegneria elettromagnetica per lo studio delle strutture guidanti è quello della trasversalizzazione delle equazioni di Maxwell. Come mai si sceglie di decomporre le equazioni di Maxwell in questa maniera? Beh, la risposta è abbastanza semplice:



Aldilà dei sofismi matematici, ciò che si vede è che esiste una direzione lungo la quale la sezione della struttura è costante: non cambia. Questo spinge a usare una scomposizione del problema di questo tipo: parte trasversale, che viene trattata con il formalismo di Marcuvitz e Schwinger fino ad arrivare a un problema agli autovalori, il quale poi può essere risolto ottenendo un'equazione di Helmholtz. Non si riconducono direttamente le equazioni di Marcuvitz e Schwinger al problema agli autovalori, ma bensì si riconduce una loro diversa rappresentazione, detta “rappresentazione modale”: si scompone l'intero campo elettromagnetico in tanti contributi, ciascuno di essi ancora soluzioni delle equazioni di Maxwell, dette **modi**. Ciascun modo, in altre parole, è una topografia di campo trasversale.

Il fatto di dividere il campo in componente trasversale e componente longitudinale ci permette di esprimere il campo elettrico in una forma del tipo:

$$\underline{E} = \underline{E}_t e^{-jkz}$$

Questo è possibile dal momento che si ha una topografia fissa rispetto alla variabile z ; in pratica si risolve il problema solo nella sezione trasversale, poi si trasporta la soluzione mediante un ritardo nel dominio dello spazio: esso deriva dal fatto che, per la sezione longitudinale, si può utilizzare il formalismo delle linee dei telegrafisti, riconducendo il calcolo del campo longitudinale ai calcoli sulle linee di trasmissione. Una guida d'onda, come il

nome suggerisce, è una struttura guidante: il campo è confinato nei pressi della struttura.

Cosa rende possibile la trasversalizzazione ? Il fatto che la sezione trasversale sia **omogenea**. Il campo \underline{E}_t è funzione delle variabili x e y , ma non è diretto solo lungo \hat{x} e \hat{y} : esistono campi trasversalmente elettrici o trasversalmente magnetici, ossia in cui il campo elettrico esiste solo nella sezione trasversale (e non in quella longitudinale), o idem solo quello magnetico: si parla di componenti TE e TM. Posso dunque spesso scrivere (nel caso vi siano tutte le componenti del campo):

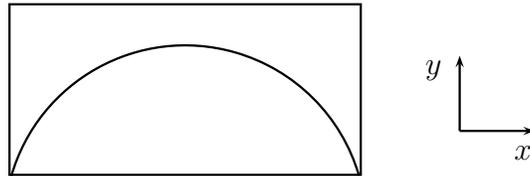
$$\underline{E} = \underline{e}_t e^{-jk_z z} + E_z \hat{z}$$

Dove uso \underline{E} come termine dal momento che essa è un fasore: la dipendenza dalla frequenza è implicita.

Il bello di aver scritto le equazioni di Maxwell nella forma trasversale è proprio il fatto che, con questo formalismo, la componente longitudinale si può ricavare banalmente dalla trasversale. \underline{e} è un'autofunzione del problema agli autovalori di cui si parlava prima. L'espansione modale viene fatta poichè semplice: in ingegneria si tende sempre a risolvere problemi mediante l'approccio *divide et impera*, e dunque questa scomposizione può essere molto interessante.

Analizziamo un caso specifico.

1.1.1 Guida d'onda rettangolare



Questa è una guida d'onda rettangolare con la sua topografia di campo fondamentale: il TE_{10} . L'andamento è un qualcosa del tipo:

$$\underline{e}_t = A \hat{y} \sin\left(\frac{\pi}{a}x\right)$$

Dunque, usando l'idea prima detta,

$$\underline{E}_{10} = A \hat{y} \sin\left(\frac{\pi}{a}x\right) e^{-jk_z z}$$

dove

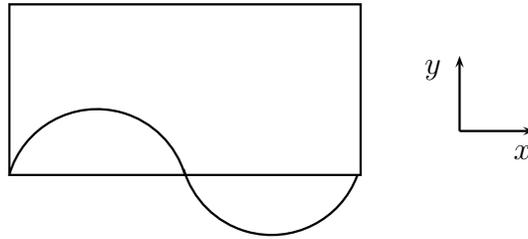
$$k_z = \sqrt{k_0^2 - k_{ti}^2}$$

e k_{ti} è:

$$k_{ti} = \frac{\pi}{a}$$

Come si ottiene matematicamente tutto ciò? Beh, si ha una guida rettangolare, dunque sul contorno della medesima si può imporre la condizione al contorno secondo cui il campo elettrico è nullo (una guida d'onda rettangolare è semplicemente un tubo rettangolare il cui contorno è fatto di metallo, e su un metallo, supponendo che sia un conduttore perfetto, il campo elettrico è nullo).

Un'altra topografia che può soddisfare la suddetta condizione al contorno però è la seguente:



Si può dimostrare in sostanza che le topografie di campo all'interno di una guida d'onda rettangolare sono sostanzialmente delle sinusoidi, del tipo

$$\sin\left(m\frac{\pi}{a}\right)$$

Al variare di m si han topografie di campo sempre diverse; interessanti per noi sono il caso per $m = 1$, e talvolta $m = 2$ (se ne parlerà in seguito).

Un'altra osservazione su k_z : il termine k_{ti} è costante, poichè dipende dalla geometria della guida rettangolare; il termine k_0 è esprimibile come

$$k_0 = \frac{2\pi f_0}{c}$$

Cosa ci dice ciò? Noi decidiamo la frequenza f_0 di lavoro; se questa è tale da avere $k_0^2 > k_{ti}^2$, allora k_z è un numero reale, e dunque il modo relativo alla k_{ti} si propaga senza attenuarsi nella struttura guidante; nel caso contrario, $k_0^2 < k_{ti}^2$, k_z è un numero puramente immaginario: questo significa che il modo relativo alla k_{ti} (l'autovalore dell'autofunzione modale) è attenuato e non si può propagare nello spazio: si propagano i modi con parte reale di k_z non nulla, si attenuano quelli con parte immaginaria non nulla. La frequenza critica dipende inoltre dalle dimensioni della guida d'onda in questione:

$$\frac{2\pi}{\lambda_{critica}} = m\frac{\pi}{a}$$

ossia

$$\lambda_{\text{critica}} = \frac{2a}{m}$$

quando $m = 1$, $a = \frac{\lambda}{2}$: questa è la frequenza di risonanza di qualsiasi roba a microonde.

Ciò che abbiamo dunque imparato è: in una guida d'onda, se la frequenza non è maggiore della frequenza critica, k_z non è un numero reale: non si ha propagazione. Se k_z è immaginario, l'esponenziale avrà forma $e^{-\alpha z}$, ossia sarà puramente attenuativo. Se $f < f_{\text{critica}}$, si parla di *guida sotto taglio*: si è sotto la frequenza in grado di far propagare il primo dei modi. Si parla di **modo fondamentale** quando si parla del primo modo che si propaga all'interno della struttura.

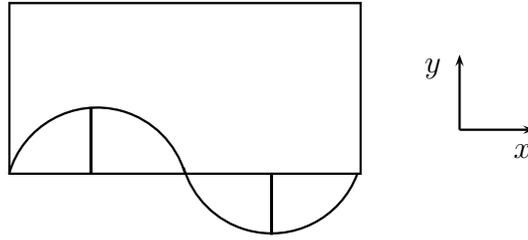
Eccitazione del modo TE_{20}

Consideriamo adesso qualcosa di particolare: di solito una struttura guidante viene utilizzata facendo propagare in essa il suo modo fondamentale, e basta. Si noti però una cosa: il fatto che un modo abbia la possibilità di propagarsi, teoricamente (praticamente poi non ce n'è tutto sto bisogno) non implica il fatto che esso esista nella struttura guidante: oltre a dover garantire di essere al di sopra della frequenza critica relativa al suddetto modo (si discuterà tra breve il caso del TE_{20}), è necessario che venga in qualche modo introdotta un'eccitazione del suddetto modo.

Per eccitare una struttura guidante è necessario introdurre delle probe, ossia delle *antennine* (di solito molto sottili e piccole), che a seconda del loro numero e della loro posizione possono eccitare uno o più modi, i quali potranno pure propagarsi, nel caso la frequenza critica relativa al modo sia superata.

Nel TE_{10} , si mette una sola probe, al centro della guida: avendo visto prima la topografia del campo TE_{10} , si vede che il punto in cui il campo è massimo è proprio quello. Siamo interessati a questa cosa (il punto in cui si massimizza il campo) dal momento che, matematicamente, una probe è così piccola (di solito) da poter essere modellata mediante una delta di Dirac nel dominio spaziale: $\delta(x - x_0)$; vale la proprietà di *campionamento* della delta di Dirac: essa "campiona" una funzione nel punto in cui essa è definita. Se il punto che facciamo campionare (essendo la funzione la famosa autofunzione) è quello massimo, allora il campo TE_{10} sarà massimizzato.

Il TE_{20} si eccita con due probe in controfase, messi a $a/4$ e $3a/4$:



Perchè tutto ciò? beh, si immagini di voler fare delle onde circolari in una pozza d'acqua: basta buttare un solo sasso in mezzo alla pozza. Per avere forme più particolari, l'eccitazione, ossia ciò che si butta nella pozza, deve essere un insieme di più sassi. Se poi la pozza è tale da permettere la propagazione di queste onde, esse si propagheranno.

Come mai si eccita così, e per esempio non così?



In questo caso la nostra sorgente è matematicamente modellabile nella seguente forma:

$$J = \delta \left(x - \frac{a}{4} \right)$$

Consideriamo a questo punto l'integrale di proiezione di questa roba: per il TE_{10} ,

$$I_{TE_{10}} = \int_0^\pi A \sin \left(\frac{\pi}{a} x \right) \delta \left(x - \frac{a}{4} \right) dx$$

questa è la componente del campo TE_{10} con questa eccitazione. Per quanto riguarda la componente TE_{20} , l'integrale è lo stesso, a meno del $m = 2$ nell'autofunzione:

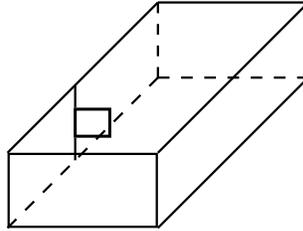
$$I_{TE_{20}} = \int_0^\pi A \sin \left(2 \frac{\pi}{a} x \right) \delta \left(x - \frac{a}{4} \right) dx$$

Cosa sono i modi? I modi sono gli assi sui quali proiettiamo i vettori, ossia i campi elettromagnetici: gli elementi della base. Gli autovalori sono quei valori che modificano solo il valore di scala, senza modificare la direzione del vettore. Fare il prodotto scalare, ossia l'integrale di proiezione, significa semplicemente calcolare le componenti, sulla base dei modi, del campo elettromagnetico.

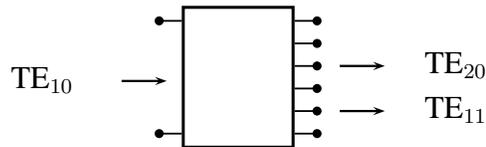
Perchè non ci piace usare una sola antennina? Perchè se ne mettiamo due, riusciamo a eccitare solo il modo TE_{20} , dal momento che si vede che si cancellano le proiezioni sul modo TE_{10} , mantenendo invece (e raddoppiando) quelle del TE_{20} grazie al fatto che si mettono in controfase.

1.1.2 Cenni alle discontinuità in guida

La presenza di discontinuità in guida d'onda crea qualcosa di interessante:



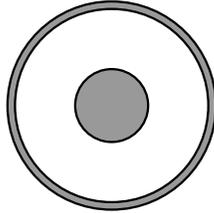
Si immagini al campo elettromagnetico come a una specie di fluido: se non vi fosse la finestrina rettangolare, il campo andrebbe dritto, senza essere fermato da nulla; quando si ha una discontinuità di questo tipo, che potrebbe derivare da una rugosità piuttosto che altro, accade che il campo va a “sbattere” contro questa finestra, e ciò provoca un’interazione tra campo e discontinuità: ciò genera un’eccitazione di tutti i modi. Si ha qualcosa di questo tipo:



Nonostante noi abbiamo un campo anche monomodale, modellando il sistema mediante matrice scattering, in uscita si hanno tutti i modi: il TE_{10} e anche i modi successivi. Quando ci sono più modi, l’unico modo per fare in maniera tale che solo un modo si propaghi, è essere sottotaglio per tutti i modi che non siano il fondamentale, e da lì i modi secondari si attenueranno senza propagarsi, con andamento esponenziale rispetto allo spazio. Una minima discontinuità fa eccitare tutti i modi, dunque basta molto poco per eccitare modi superiori; per questo motivo è fondamentale lavorare a frequenze corrette, frequenze di monomodalità. Questo è vero nel mondo reale: nella realtà tutto ciò che può essere eccitato è autoeccitato; per questo motivo è necessario lavorare sottotaglio.

1.1.3 Cavo coassiale

Senza voler risolvere le equazioni di Marcuvitz e Schwinger / Helmholtz per questo tipo di problemi (bisognerebbe passare in coordinate polari per la sezione trasversale e si otterrebbero le equazioni di Bessel, le cui soluzioni sono le note funzioni di Bessel), si ha qualcosa di diverso:



Anche il cavo coassiale ha una frequenza di taglio inferiore per il primo modo, per il modo fondamentale: 0 Hz. Questo dal momento che il cavo coassiale è una struttura in grado di supportare un modo TEM, il cui autovalore è nullo: in un coassiale è possibile trasportare anche la continua.

Le caratteristiche geometriche in ogni caso sono quelle che fanno variare la frequenza superiore, ossia la frequenza di propagazione del primo modo: a seconda del diametro esterno (discutiamo ora la cosa), è possibile modificare le caratteristiche della linea equivalente.

Si ricordi la regola fondamentale: più le dimensioni fisiche (in questo caso il diametro esterno) sono ridotte, più la frequenza di taglio superiore sarà alta: per questo motivo, a microonde, più si sale di frequenza e più le cose devono essere **piccole**. Un modo mnemonico è: più si aumentano i constraints, più si vincola (riducendo le dimensioni fisiche) il campo, più i modi saranno distanziati tra loro: se nello spazio libero lo spettro dei modi è continuo, dal momento che non si hanno sostanzialmente dei vincoli, più si limita e più i modi si distanziano. La frequenza di taglio andrà poi a limitare l'applicazione, dunque la struttura guidante andrà progettata basandosi su questo principio.

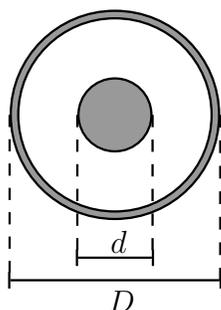
I coassiali sono fondamentali soprattutto dal momento che spesso si utilizzano connettori a microonde: se usiamo un connettore da 100 MHz a 10 GHz, capita che si ha un'enorme attenuazione del segnale: il connettore eccita tutti i modi, questi vengono attenuati dal cavo con un'attenuazione molto elevata, e per questo si ha molta attenuazione.

Finora si è parlato di condizioni al contorno per quanto riguarda il campo trasversale; in realtà è necessario introdurre condizioni al contorno anche per quanto riguarda il campo elettromagnetico longitudinale: dal momento che il modello che si utilizza per la propagazione longitudinale è quello delle linee di

trasmissione, dunque un modello circuitale, le condizioni al contorno saranno introdotte introducendo impedenze di carico e di sorgente. Si definisce, per le linee di trasmissione, un'impedenza di riferimento Z_∞ : essa è quell'impedenza di carico tale per cui, se introdotta, fa comportare la linea di trasmissione come una linea di lunghezza infinita: una linea che non ha riflessione. Lunghezza infinita è il significato del pedice ∞ . Storicamente si usa, come impedenza caratteristica, $Z_\infty = 50\Omega$.

Per quanto riguarda il cavo coassiale, vale la seguente formula:

$$Z_\infty = Z_0 \ln \left(\frac{D}{d} \right)$$



Questa formula è giustificabile mediante un semplice ragionamento statico: se ho che D e d sono uguali, ossia che le due parti di metallo si toccano, l'impedenza sarà nulla; più si distanziano, più aumenta l'impedenza. Più aumenta D/d , più l'impedenza aumenta, per questo si fa così; se $D = d$, il logaritmo di 1 è 0.

Quanto vale Z_0 ? Beh, si ricordi che uno dei connettori principali è “il 7 mm”: il diametro esterno è di 7 mm, quello interno 3 mm, connettore a 50Ω . Si ha:

$$50\Omega = \ln \left(\frac{7}{3} \right) Z_0$$

da qui

$$Z_0 = 59\Omega$$

spesso si usa $Z_0 = 60\Omega$. Questa condizione impone il fatto che l'impedenza del connettore sia 50Ω . Ciò ci farà capire al volo qual'è l'impedenza dei cavi coassiali date le dimensioni fisiche.

Andiamo avanti: la legge di Ohm statica e banale afferma che la resistenza R di qualcosa può essere calcolata come:

$$R = \varrho \frac{l}{A}$$

Ho un conduttore con un diametro centrale di 3 mm, il campo maggiore sarà vicino al conduttore centrale (essendo il potenziale applicato su di esso), ma dunque più piccolo è il conduttore centrale, maggiori saranno le perdite: riducendo il diametro riduco la sezione, ma dunque aumento (qualitativamente parlando) la resistenza, ma con essa dunque anche le perdite. Avere un conduttore centrale piccolo eppure è necessario: la nostra necessità è quella di aumentare la frequenza di taglio del primo modo superiore, dunque per farlo, come già detto in precedenza, è necessario proprio ridurre le dimensioni dei conduttori.

Al crescere della frequenza, si avrà un aumento delle perdite; tipicamente, questo aumento va come \sqrt{f} . Il nostro obiettivo è:

- mantenere la giusta impedenza;
- evitare i modi surmodati.

Si avrà un aumento non lineare delle perdite: A va col quadrato della lunghezza, per questo le perdite avranno un andamento non lineare (qualitativamente: sono ragionamenti statici, che però sono qualitativamente validi). Questo concetto è vero anche in guida d'onda, anche se molto meno preoccupante.

Capitolo 2

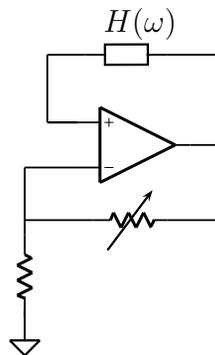
Generatori a microonde

2.1 Concetti fondamentali per generare segnali a microonde

Al fine di generare segnali a microonde, è necessario capire quali componenti siano necessari per farlo. Fondamentale, soprattutto, sarà il ruolo dell'oscillatore, Come si potrebbe fare, per realizzare un oscillatore a bassa frequenza? Sostanzialmente sono necessari due componenti:

- una parte amplificatrice;
- un feedback positivo.

L'idea potrebbe per esempio essere la seguente:



La funzione di trasferimento del filtro $H(\omega)$ deve essere tale da far oscillare a una frequenza particolare il circuito, attenuando le altre; la parte di amplificazione è necessaria dal momento che, se non limitato il guadagno dell'amplificatore (potrebbe per esempio essere un operazionale), esso potrebbe

saturare, producendo un'onda quadra; parlando di oscillatori per ora siamo interessati a oscillatori di tipo **sinusoidale**, dunque servirà anche una sorta di reazione negativa.

Il fatto di trovarci a microonde porta alcuni vantaggi e molti svantaggi:

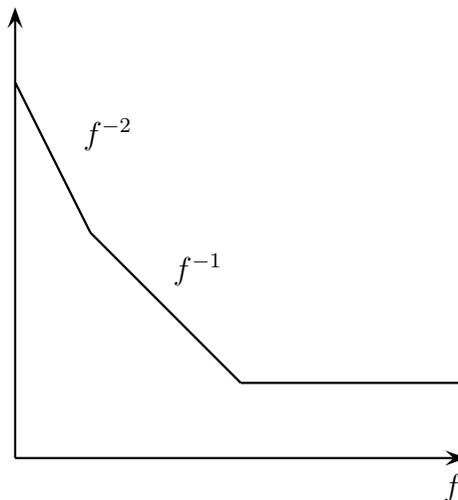
- uno degli unici vantaggi a nostro favore è il fatto che possiamo limitare facilmente le armoniche: dal momento che siamo ad alte frequenze, le seconde armoniche che si vengono a formare sono a frequenze altissime, dunque il filtraggio non è complicato;
- uno dei grossi svantaggi sta nel fatto che avere un elemento equivalente al quarzo è molto difficile, dunque ci si dovrà arrangiare con altri mezzi.

Passiamo a questo punto a un'analisi delle caratteristiche di base di un oscillatore, nella fattispecie a microonde.

2.1.1 Purezza spettrale: phase noise

Una delle prime caratteristiche è la purezza spettrale. Si consideri una $H(\omega)$ del tipo:

Come detto l'oscillatore in questione è sinusoidale; nel dominio di Fourier questo significa che si dovrebbe avere a che fare con una risposta del tipo impulsivo: lo spettro dovrebbe essere una singola riga. Nella pratica non è così, dal momento che essa presenta un andamento a campana; al fine di quantificare il phase noise ciò che si fa è calcolare la potenza di rumore in una banda di 1 Hz a una certa distanza dalla carrier (portante, alla frequenza di risonanza). Si trova, in funzione di f , qualcosa del tipo:

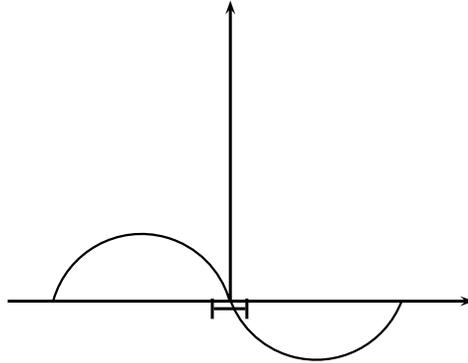


Si noti una cosa: idealmente si potrebbe vedere ciò con l'analizzatore di spettro ma, in pratica, sono necessari alcuni trucchi.

Il grafico del phase noise (campana) mostra che vi sono dei termini spuri, dovuti a una modulazione FM residua e a battimenti che si formano all'interno del sintetizzatore.

Esistono diversi modi di ottenere oscillatori a una certa frequenza: sostanzialmente due vie possono essere essere quella di fare una **sintesi diretta a microonde** piuttosto che una **sintesi per moltiplicazione**, a partire da una frequenza bassa. La seconda tecnica a occhio è più semplice da realizzare, ma come mai non si fa? La risposta è nella seguente figura:

Queste sono le rappresentazioni degli andamenti del phase noise nel caso in cui si ottenga il segnale per sintesi diretta o per moltiplicazione; moltiplicando in frequenza, il phase noise aumenta sempre. Come mai? Beh, il phase noise ha ovviamente significato anche nel dominio del tempo: si tratta di un'incertezza nel tempo di transito, ossia del **jitter**:



Nel caso dei sistemi digitali, per esempio, il jitter del clock deriva proprio dal phase noise dell'oscillatore. Come mai esso è presente? Beh, abbastanza semplice: un oscillatore, per quanto valido, ha una determinata incertezza; questa si può vedere, nel dominio del tempo, come un jitter, un errore sull'attraversamento in un periodo di segnale; nel caso in cui si ottenga la frequenza per moltiplicazione, si chiede che lo stesso segnale, con le proprie incertezze, sia ripetuto più volte nello stesso tempo, cosa che di fatto moltiplica, per il fattore di moltiplicazione, anche le incertezze, i jitter, rendendo peggiore il rumore di fase.

2.1.2 Altre caratteristiche

Il phase noise è solo una delle caratteristiche degli oscillatori: se quelle che ci hanno per ora infastidito sono impurità spettrali causate da errori di fase,

in realtà le impurità stesse hanno più cause: potrei avere problemi causati da una modulazione AM residua. Il rumore di fase dovuto all'oscillatore è più importante del rumore di ampiezza, e più difficile da gestire, dunque di solito si cerca di fare più attenzione a questo. Fatto sta, per quanto riguarda questo parametro, che è fondamentale anche **stabilizzare l'ampiezza**.

Un secondo problema potrebbe essere il fatto che il sistema può essere sensibile alle variazioni della tensione di alimentazione V_{AL} , della temperatura o di altro; si tratta generalmente di parametri di secondaria importanza, dunque non vi spendiamo troppe parole, come per le modulazioni residue AM.

2.1.3 Circuiti risonanti

Il nostro obiettivo è quello di realizzare un oscillatore a microonde, spettralmente puro, con banda di frequenza variabile (al fine di poter fare degli sweep di frequenza, e poterli usare nella strumentazione a microonde). Per fare ciò sarebbe necessaria una $H(\omega)$ variabile, controllabile, in cui la frequenza di risonanza sia modificabile in qualche maniera; dal momento che H è la funzione di trasferimento di un filtro, ciò che ci servirebbe è un filtro variabile, una rete di reazione positiva variabile.

Per quanto riguarda dunque i filtri variabili, riprendiamo brevemente un concetto derivante dalla teoria dei circuiti: la purezza spettrale dipende sostanzialmente dall'ampiezza della campana, nel dominio della frequenza, del filtro: più la campana è stretta, migliore sarà la purezza spettrale. Si può vedere la cosa sotto questo punto di vista: se l'amplificatore, l'altro componente fondamentale del sistema, è una sorta di generatore di rumore bianco, la H ha lo scopo di filtrare questo rumore bianco; il fatto che l'andamento in frequenza sia a campana comporta che anche l'andamento del rumore filtrato sia a campana, ma dunque che più il filtro è stretto, maggiore sarà il filtraggio.

Ciò che quantifica la qualità del filtro è il cosiddetto **fattore di qualità**, Q . Si consideri un circuito risonante del tipo:



Nel caso in cui non si abbia alcuna resistenza, ossia in cui non si abbiano perdite di tipo ohmico, $Q \rightarrow \infty$; questo significa che, in quel caso, la campana degenera sostanzialmente nella tanto desiderata δ di Dirac. Si ha infatti che:

$$Q = \omega_0 RC$$

dove, nel caso di risonatore ideale (senza perdite), $R \rightarrow \infty$. Vale:

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

Definizione alternativa è quella che coinvolge la frequenza di risonanza f_0 e la larghezza di banda a -3 dB intorno a essa, Δf :

$$Q = \frac{f_0}{\Delta f}$$

Più la campana è stretta, più Q è elevato, dunque Δf è a denominatore.

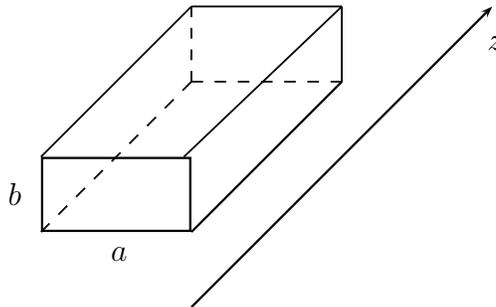
Esiste una terza definizione, collegata al significato energetico del Q ; essa può essere ricavata secondo il seguente ragionamento: il Q è collegato alle perdite ohmiche del circuito, nel senso che maggiori sono le perdite, maggiore è la resistenza, minore è Q ; parlare di perdite ohmiche si traduce nel parlare di potenza attiva P_A : maggiore sarà la potenza dissipata, minore sarà il Q ; dualmente, maggiore sarà la potenza reattiva rispetto a quella dissipata, maggiore sarà Q . Si può dunque ricavare da questo ragionamento:

$$Q = \frac{P_{\text{reattiva}}}{P_{\text{attiva}}}$$

2.2 Realizzazioni di oscillatori

2.2.1 Oscillatore basato su cavità risonante

Una cavità risonante è semplicemente pensabile come una guida d'onda, in cui si chiude a un certo punto la sezione trasversale:



Come mai siamo interessati a qualcosa di questo tipo? Beh, si pensi al nome: cavità **risonante**: all'interno della cavità è possibile generare dei

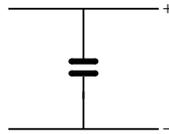
modi che continuano a risonare all'interno di essa. Di fatto, come avevamo detto, la guida d'onda è modellabile mediante una linea di trasmissione; in questo caso, ciò che si farebbe è chiudere la suddetta guida d'onda su due corti, sia a sinistra sia a destra. Per una qualche frequenza, si avrà che la rete sarà chiusa per una certa $\lambda_g/2$, legata alla frequenza e alla geometria della struttura; dal momento che, come accennato, *tutto risuona a $\lambda/2$* , variando la geometria della guida, cambiando la distanza tra le varie pareti, si cambia $\lambda_g/2$ e con essa la frequenza di risonanza della struttura.

Tanto più la struttura guidante è senza perdite, tanto migliore sarà il Q del risonatore ottenuto a partire dalla cavità; con le cavità si possono ottenere Q molto elevati (migliaia e migliaia), dal momento che essi sono sostanzialmente condizionati dalla conducibilità del materiale conduttore costituente la parete della struttura guidante; muovendo meccanicamente la parete con delle viti, dunque, è possibile modificare la risonanza.

Uno dei vari problemi di questa struttura sta nel fatto che essa è *tutta una parete*: di fatto di modi fondamentali che potrebbero essere eccitati ve ne sono diversi, dal momento che si hanno modi a tre indici (tre confinamenti spaziali del contorno portano alla discretizzazione anche della terza dimensione); capire come eccitare la cavità in maniera tale da ottenere proprio il modo desiderato è complicato.

2.2.2 Varactor - richiamo al calcolo della sensibilità

Modificare le pareti della cavità risonante è sostanzialmente riconducibile, pensando a parametri concentrati, alla variazione degli elementi reattivi presenti nel sistema. Qualcosa di questo tipo si può fare anche elettricamente (invece che meccanicamente), con dispositivi allo stato solido: i **varactor**. Si tratta sostanzialmente di diodi dei quali però siamo interessati a un particolare parametro: la capacità di svuotamento (ossia la capacità di giunzione inversamente polarizzata). Si ha un comportamento di questo genere:



$$C_{\text{varactor}}(V) = \frac{C_p}{\left(1 - \frac{V}{\phi}\right)^m}$$

dove di solito $m = \frac{1}{2}$.

Questa cosa funziona, e viene talvolta utilizzata, ma ovviamente come ogni cosa ha dei problemi: la capacità, sebbene il modello come la radice quadrata sembrerebbe suggerire enormi variazioni di capacità, non varia più del 10%. Con una variazione di questa portata, quanto può variare la frequenza in un circuito risonante? Vediamolo: il nostro obiettivo è trovare quali sono le variazioni di frequenza relative, causate da una variazione di capacità; in matematica:

$$\frac{\Delta f}{f} = \frac{\Delta C}{C}$$

Dove per Δf si intende l'**incremento finito** di f . Utilizzando l'approssimazione di primo grado mediante derivata, si può dire che l'incremento finito Δf dovuto a un incremento finito di C , ΔC , sia:

$$\Delta f = \frac{\partial f}{\partial C} \Delta C$$

Ricordo a questo punto che:

$$f = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{L}} C^{-\frac{1}{2}}$$

dunque:

$$\frac{\partial f}{\partial C} = -\frac{1}{2} C^{-\frac{3}{2}} \frac{1}{2\pi\sqrt{L}}$$

A questo punto, per considerare le incertezze relative:

$$\begin{aligned} \frac{\Delta f}{f} = \frac{\Delta C}{C} &\Rightarrow -\frac{1}{2} \frac{C}{C^{3/2}} \frac{1}{2\pi\sqrt{L}} = \\ &= \frac{1}{2} \Delta C \end{aligned}$$

Ossia, se la variazione possibile di ΔC è del 10%, la massima variazione da essa derivante di Δf sarà del 5 %: assolutamente irrisoria. Questo ci dice che il metodo utilizzato funziona, ma di sicuro non può essere usato per generatori di segnale a larga banda.

2.2.3 YIG: Yttrium Iron Garnet

Uno degli oscillatori più interessanti è stato scoperto negli anni '50 da un premio Nobel: si tratta di oscillatori basati su materiali ottenuti per sinterizzazione (come il gesso: produzione ad alta temperatura e alta pressione).

Uno YIG è una piccola sfera con una caratteristica molto particolare: è un materiale ferromagnetico con costante μ elevata, in modo che, se immerso in un campo magnetico **statico**, si comporta da filtro a microonde, con un fattore di qualità Q elevatissimo.

Come funziona questa cosa? Si immagina che si abbia una piccola spira lungo \hat{x} , un'altra spira ad essa ortogonale lungo \hat{y} , e dunque il campo magnetico statico lungo l'asse dei magneti, ossia lungo \hat{z} ; si introduce questo sistema all'interno di un solenoide (in un campo magnetico statico). Un materiale ferromagnetico ha una costante μ che dipende dal campo magnetico \mathcal{H} ; per le equazioni di Maxwell, si ha che il campo elettrico va a dipende dal campo magnetico **statico** (dinamico è normale), ossia alla continua. L'andamento *a campana* di questo sistema si sposta al variare del campo magnetico statico lungo \hat{z} . La forma a campana è dovuta al fatto che il segnale passa solo quando si ha un effetto di risonanza della pallina di YIG: le due spire non sono a contatto tra loro, dunque è lo YIG a stabilire quando il segnale debba passare: una sorta di trasformatore con traferro risonante, la cui frequenza di risonanza è variabile.

Il bello degli YIG è il fatto che la campana, tra la frequenza minima e la frequenza massima, si allarga molto poco: alzando la frequenza di risonanza dunque, tendenzialmente, il Q aumenta; sicuramente un minimo di allargamento vi sarà, ma non troppo evidente. Facciamo due conti sul Q : dei numeri ragionevoli sono 20 MHz di banda a 20 GHz; si ha qualcosa tipo

$$Q = \frac{20000}{20} = 1000$$

In un risonatore in microstriscia, per esempio, $Q = 15 \div 20$; in guida, si parla di parecchie migliaia.

Nei telefoni cellulari in realtà non si usa nulla di tutto ciò: al fine di realizzare filtri si usano filtri SAW (Surface Acoustic Waves); molti pensano che il componente che ha permesso di *fare il telefonino* sia stato il transistor; nella realtà ciò che ha veramente dato la spinta è stato il poter fare filtri senza DRO o cavità o altro, ma SAW: Q elevati anche a microonde e dimensioni ridotte.

Abbiamo finora parlato bene degli YIG, ma qual è il loro problema? Beh, abbiamo detto che la loro frequenza di risonanza si controlla controllando il campo magnetico statico; un campo magnetico statico, come noto dalla Fisica, si può produrre con un solenoide o comunque semplicemente facendo correre della corrente in un filo idoneo. Generalmente, uno YIG ha una sensitivity di 20 MHz/mA : per ogni mA si ha una variazione di 20 MHz. Per far oscillare lo YIG a 20 GHz, per esempio, serve una corrente di 1 A.

Osserviamo ancora una cosa dei circuiti risonanti:

Il fatto di aver Q elevato comporta un altro fatto, oltre a ciò che è stato già detto: più Q è elevato, tanto più ripida sarà la transizione di fase. Il nostro obiettivo è quello di avere una transizione di fase molto larga: più si ha una fase ripida, più eventuali variazioni di fase si traducono in una variazione di frequenza, piccola.

Di solito uno YIG varia di 200° in 20 MHz; questo significa che, con 1 mA, spostiamo di 200° . Tradotto al grado,

$$\frac{1\text{mA}}{200^\circ} = 5\mu\text{A}/^\circ$$

Questi sono parametri tipici.

Cosa significa ciò? Si deve progettare un oggetto con un campo magnetico statico e una corrente di controllo che deve stare ferma di $5\mu\text{A}$ (per ammettere una deviazione standard di 1°). Se c'è una variazione di fase di 1° relativa a una qualche variazione di campo magnetico, qual è la variazione di frequenza? Semplicemente, con la proporzione:

$$1 : 360^\circ = \Delta f : f_{\text{centrale}}$$

Si supponga $f_{\text{centrale}} = 3,6\text{ GHz}$; con 1° di variazione ho:

$$\Delta f = \frac{f_{\text{centrale}}}{360^\circ} 1^\circ = 10\text{MHz}$$

Ossia, con 1° di deviazione standard, abbiamo 10 MHz di deviazione standard. Per avere, con 20 MHz/mA di variazione, 3,6 GHz con precisione 10 MHz, dobbiamo avere 180 mA di corrente fissa, e questa deve essere esatta con un errore di $5\mu\text{A}$: questo è difficilissimo da realizzare, dal momento che questa corrente deve essere precisissima (errore del 0,0027 %). Il $\Delta f/f$ si traduce immediatamente in un controllo del campo magnetico statico, dunque questo controllo deve essere molto efficace. Esiste un secondo problema negli YIG: si deve fare in modo che il campo statico non venga influenzato da campi esterni, quali quelli di magneti o derivanti da altro; si ha inoltre una certa dipendenza anche dalla temperatura.

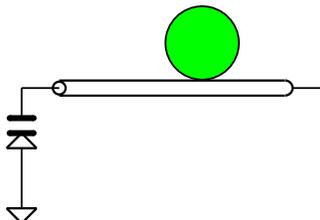
Far funzionare sul serio un circuito di controllo degli YIG è una cosa assolutamente non banale da fare.

2.2.4 DRO: Dielectric Resonators Oscillators

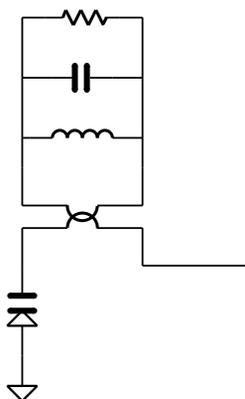
Se l'obiettivo del progetto è quello di realizzare oscillatori a frequenza fissa, si possono utilizzare i DRO: Dielectric Resonator Oscillators. Si tratta anche in questo caso di oscillatori basati su materiali ottenuti per sinterizzazione,

ma in questo caso dalle proprietà dielettriche interessanti: si parla di cavità dielettriche, in cui la ϵ_r è elevatissima. Anche in questo caso, si ha a che fare con fattori di qualità Q molto elevati, dal momento che le perdite sono bassissime.

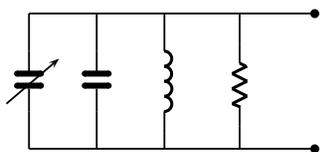
Per accoppiare il risonatore al resto del circuito si fa qualcosa di questo genere:



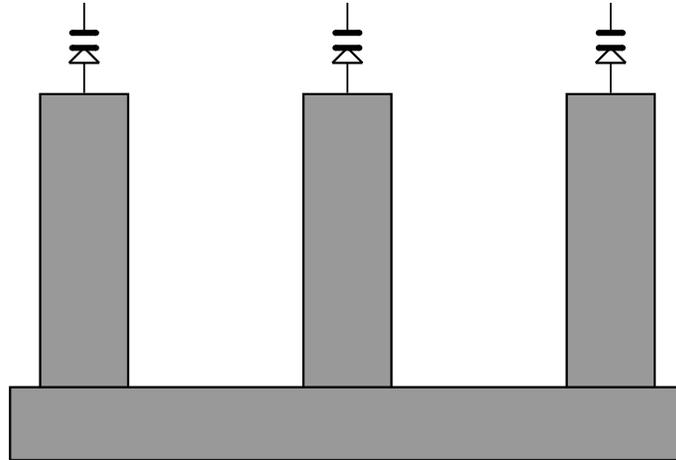
Il circuito elettrico equivalente di questo sistema è qualcosa di questo genere:



Si ha un accoppiamento di questo genere, modellabile mediante trasformatore induttivo; la pastiglia si comporterà da risonatore. Si usa aggiungere un varactor, al fine di far variare la capacità equivalente del circuito DRO + microstriscia; si ottiene, in risultato finale, qualcosa del tipo:



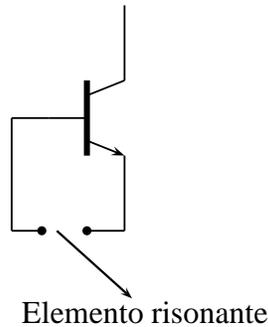
Una soluzione proposta per ottenere una funzione di trasferimento un poco più articolata potrebbe essere la seguente:



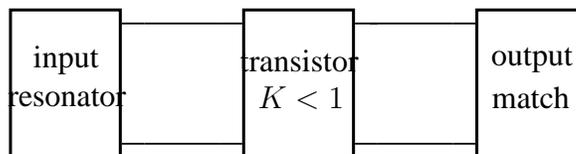
Si hanno tre stub in microstriscia dove ciascuno stub termina su un varactor; in questa maniera si riesce ad avere tre circuiti risonanti la cui frequenza di risonanza è leggermente modificabile mediante il suddetto varactor. La funzione di trasferimento sarà dunque più complicata di quelle finora viste.

2.2.5 Stadio di amplificazione

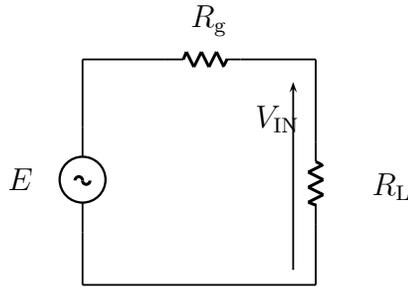
Come si può realizzare, a microonde, l'amplificatore atto a evitare la saturazione? Ci sono diverse soluzioni, per esempio accoppiare con un risonatore base ed emettitore:



Ciò che nella realtà si fa nella maggior parte dei casi a microonde è utilizzare la *tecnica per riflessione*: se il coefficiente di stabilità K è minore di 1, dunque il transistor instabile, il coefficiente di riflessione all'ingresso, Γ_{ingresso} , è fuori dalla carta di Smith (dal momento che ha modulo maggiore di 1). Si ha una situazione di questo genere:



Cerchiamo di capire come funziona a partire da una situazione tipica in bassa frequenza: a bassa frequenza si potrebbe avere qualcosa del tipo:



dove

$$V_{IN} = \frac{R_L}{R_g + R_L} E$$

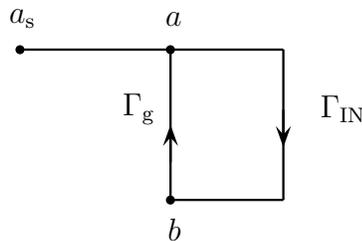
L'obiettivo, per avere oscillazioni a partire da un circuito equivalente di questo tipo, è avere il denominatore a 0, e $E = 0$: in questa maniera si ha un segnale che continua a girare nel circuito senza la presenza di un'eccitazione.

Nontouching loop rule

A microonde si fanno dei ragionamenti, invece che su tensioni e correnti, su onde e coefficienti di riflessione; ciò che si fa è utilizzare i **grafi di Mason**: si scrive l'onda incidente su ciascun nodo (e il nodo dunque indica l'onda), dunque quando c'è il parametro di riferimento metto $\Gamma_{\text{riferimento}}$, nell'altro Γ_g . Si fa in questo modo la somma di ciò che arriva sul nodo:

$$b = \Gamma_{in} a$$

$$a = \Gamma_g b + a_s$$



Come mai si fa uso di questi grafi? Perchè esiste una regola, sempre di Mason, che permette di calcolare la funzione di trasferimento di un grafo: la

nontouching loop rule, derivante da un articolo del '55 per l'appunto di Mason. Questa regola dice che, guardando i loops, devo guardare, per avere una certa funzione di trasferimento, qual è il percorso che collega direttamente l'uscita all'ingresso; infine, si divide questo percorso per tutti i loops che non toccano l'ingresso. Si consideri per esempio la funzione di trasferimento a/a_s : tra a_s e a il rapporto è unitario, dunque divido per l'unico loop che non tocca a_s : $1 - \Gamma_g \Gamma_{in}$, ottenendo:

$$a = a_s \frac{1}{1 - \Gamma_g \Gamma_{in}}$$

Cosa ci dice questa formula? Beh, Γ_g sarà sicuramente in modulo minore o uguale a 1, essendo esso derivante da un'impedenza con parte reale positiva; i circuiti risonanti che possono infatti portare alla definizione di coefficiente di riflessione di generatore sono tutti circuiti passivi o nel caso migliore reattivi. Quello che serve, per avere l'effetto che desideriamo (analogo all'effetto in bassa frequenza), è $a_s = 0$, e Γ_{in} maggiore di 1, in modo da annullare il denominatore.

Si vuole a questo punto proporre la regola in modo pi formale, al fine di renderla applicabile a qualsiasi esempio.

- volendo scrivere una generica funzione di trasferimento, prima di tutto si scriva il denominatore della medesima: esso deve contenere il termine additivo 1, e vi devono essere sommati tutti i possibili loop che non si toccano; nel caso dei loop di primo ordine, ossia di quelli costituiti da un singolo loop, si pu avere in comune un nodo, ma non un ramo; a seconda dell'ordine del loop, ossia, per loop "composti" da loop non toccanti, il segno sar + nel caso in cui l'ordine del loop (il numero di loop che moltiplichiamo assieme) pari, - se viceversa l'ordine dispari. Nel caso della figura precedente dunque si avr, a denominatore:

$$H = \frac{\Xi}{1 - T_1 - T_2 - T_3 + T_1 T_3} = \frac{\Xi}{1 - h - fg - de + hde}$$

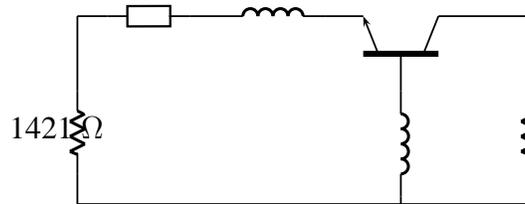
- al numeratore, si sommano i guadagni, ciascuno dei quali moltiplicato per $1 \pm \xi$, dove per ξ si intende in questo caso la somma/differenza dei **guadagni di anello dei loop che non toccano n per archi n per nodi il cammino del guadagno**, e il segno si regola di nuovo come prima. In questo caso:

$$H = \frac{G_1 \times (1 - T_1 - T_2) + G_2(1 - T_1)}{1 - T_1 - T_2 - T_3 + T_1 T_3} = \frac{G_1(1 - h) + G_2(1 - h - fg)}{1 - h - fg - de + hde}$$

infatti, come si pu vedere dalla figura, il guadagno G_1 tocca solo per un nodo il loop T_3 , ma non gli altri; al contrario, G_2 tocca T_2 e T_3 .

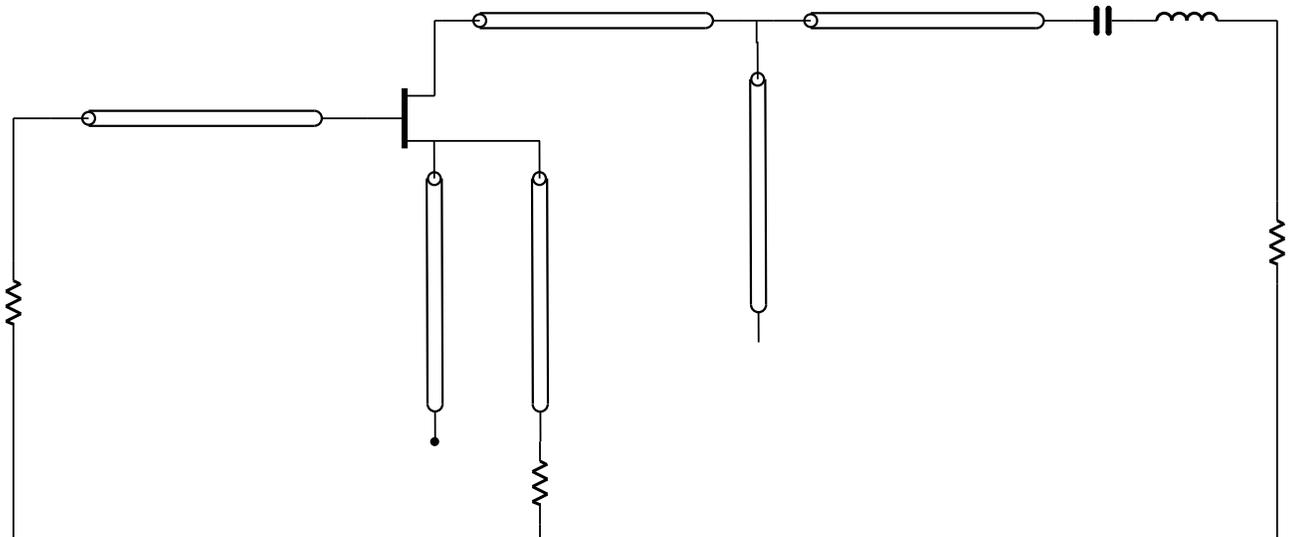
Realizzazione mediante oscillatori a diodo Gunn

Come si può realizzare una cosa del genere? Un esempio di schema potrebbe essere il seguente:



Si tratta sostanzialmente di uno stadio a base comune, e sull'emettitore si sostituisce quello che faceva da DR con una impedenza molto elevata e linea di trasmissione con ϵ_r molto elevata: il tratto di linea di fatto si comporta come un risonatore (in questo caso come un risonatore serie), e quindi si ha un risonatore. L'obiettivo è quello di far vedere un'impedenza negativa, ma ciò si ottiene grazie alla presenza dell'induttanza.

Una cosa più classica e più utilizzata potrebbe essere la seguente:



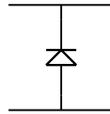
In questo caso si hanno quasi solo linee di trasmissione; il DR viene schematizzato con linee di trasmissione più resistenza; conosco la frequenza di risonanza del DR, e il DR alla risonanza si comporta sostanzialmente

come una resistenza. Se si fa in modo tale che la resistenza venga riportata con la linea di trasmissione in fase con il Γ_{in} generato in ingresso al transistor, il problema è risolto. Questo si può fare con calcoli di Campi Elettromagnetici, anche senza simulatori: si mettono un po' di induttanze sul drain, al fine di introdurre un'instabilità elevata; il DR alla frequenza di risonanza si comporta come una resistenza, si fa in modo tale che questa resistenza sia riportata alla frequenza di oscillazione con la fase opportuna rispetto al Γ_{in} del transistor, e il gioco è fatto.

2.2.6 Oscillatori a diodo Gunn

Un'altro metodo per realizzare oscillatori è basato sui cosiddetti “diodi Gunn”. Si tratta di diodi, realizzati su substrati veloci (come il GaAs), che hanno una caratteristica tensione/corrente di questo tipo:

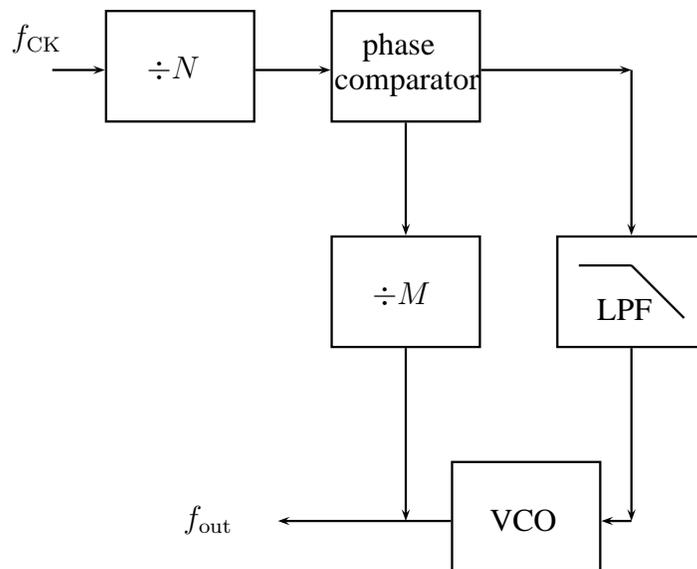
Nel caso del diodo normale, la caratteristica è una funzione ed è monotona crescente; nel caso del diodo Gunn si ha un punto di polarizzazione con resistenza differenziale negativa, ossia in cui la pendenza della retta tangente è per l'appunto negativa. Nel caso in cui si polarizzi il diodo Gunn in uno di questi punti, esso genererà un coefficiente di riflessione Γ tale per cui $|\Gamma| > 1$. Ciò si può sfruttare nella seguente maniera:



Se si introduce in questo modo il diodo Gunn all'interno di una guida, esso può compensare le perdite del risonatore, permettendo dunque ad esso di oscillare.

2.2.7 Sintetizzatori a microonde

Si è detto che la purezza spettrale dell'oscillatore dipende dal fattore di qualità, Q , del risonatore. Come si può agire, per avere un'elevata purezza spettrale? Sostanzialmente, ciò che si fa è agganciare l'oscillatore a microonde a un certo riferimento; questo riferimento è un sintetizzatore a microonde, realizzato mediante un PLL, ossia mediante un Phase Lock Loop. Lo schema di principio di un sintetizzatore a PLL è il seguente:



Si ha

$$\frac{f_{\text{out}}}{M} = \frac{f_{\text{CK}}}{N}$$

dunque

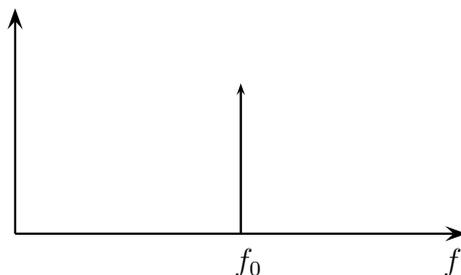
$$f_{\text{out}} = \frac{M}{N} f_{\text{CK}}$$

Qual è l'elemento critico di questo sistema? I divisori di frequenza: il segnale di riferimento potrebbe essere a qualche centinaio di MHz, dunque di fatto gli altri elementi potrebbero non essere particolarmente complicati da realizzare. Il segnale in uscita dal VCO potrebbe avere una frequenza f_{VCO} anche pari a 20 GHz, dunque altissima. I divisori di frequenza sono blocchi sostanzialmente digitali: sono dei prescaler, ossia divisori per potenze di 2; questi si realizzano, in pratica, mediante degli *shift register*. Al fine di realizzare questi prescaler, dunque, dovremo sostanzialmente fare una cascata di flip-flop, in maniera da realizzare divisori per 2^N . Il problema sta nel fatto che il digitale, per funzionare a microonde, deve avere alle spalle un processamento del segnale a microonde: il segnale digitale deve almeno assomigliare a un'onda quadra, alle microonde, e deve avere dunque un mucchio di armoniche per poter ricordare appunto un'onda quadra (di solito si richiedono almeno 10 armoniche); questo richiede che ogni singolo transistor nel circuito abbia una frequenza f_T superiore almeno ai 100 GHz (che, con un segnale a 20 GHz, sarebbe a malapena la quinta armonica, dunque servirebbe molto più alta). Già è un problema entrare nel circuito con una

forma d'onda quadra, dal momento che l'oscillatore oscilla in sinusoidale; il primo transistor lo deve convertire in una quasi-onda quadra, tutti gli altri andran dietro a esso; tutte le f_T dei transistori dovranno essere elevatissime.

2.2.8 Divisione per sottocampionamento

Si consideri un segnale sinusoidale alle microonde:



Se io campiono questo segnale con un campionatore ideale (che dunque campiona esattamente al momento desiderato e esattamente il valore desiderato), normalmente per mantenere l'informazione nel segnale originario è necessario campionare con frequenza di campionamento f_C che sia almeno doppia rispetto a f_0 , per rispettare il teorema di Nyquist e dunque poter ricostruire la forma d'onda.

Si noti però un fatto interessante: in un segnale sinusoidale, di fatto, non vi è informazione: noi sappiamo come è fatto un segnale sinusoidale una volta definite ampiezza, fase iniziale e frequenza; a parte ciò, non vi sono altre informazioni, in un segnale di questo tipo. Questa cosa suggerisce che, di fatto, non è necessario mantenere le informazioni portate dalla carrier o da segnali a frequenze elevatissime, bensì solo mantenere le informazioni del segnale utile; per questo motivo si parla di **sub-Nyquist sampling**, o di **campionamento a tempo equivalente**.

Come si fa? Cosa significa tutto ciò? Beh, ricordiamo come opera un campionatore ideale: esso semplicemente moltiplica, nel dominio del tempo, la forma d'onda interessata per un treno di impulsi, ossia per un treno di δ di Dirac. Ricordando, da Teoria dei Segnali, che la trasformata di un treno di impulsi è ancora una volta un treno di impulsi, si avrà:

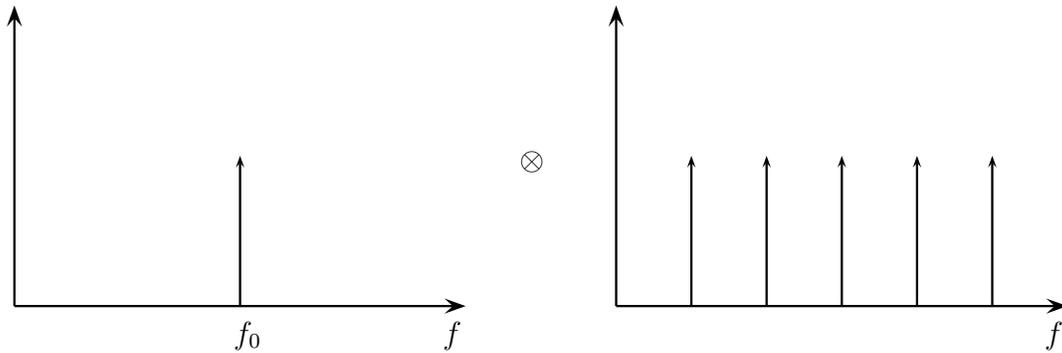
$$\sum_{n=-\infty}^{+\infty} \delta(t - nt_{CK}) \xrightarrow{\mathcal{F}} \sum_{n=-\infty}^{+\infty} \delta(\omega - n\omega_{CK})$$

Cosa significa ciò? Beh, si consideri un segnale $x(t)$: campionandolo, si ha:

$$\sum_{n=-\infty}^{+\infty} x(t)\delta(t - nt_{\text{CK}}) \xrightarrow{\mathcal{F}} \sum_{n=-\infty}^{+\infty} X(\omega) \otimes \delta(\omega - n\omega_{\text{CK}})$$

dove \otimes è l'operatore di convoluzione. Questo significa che il problema del campionamento si deve semplicemente ricondurre al problema della trasformazione di Fourier del pettine, ossia $X(\omega)$. Questa è, generalmente, la parte difficile del problema in questione.

La convoluzione di delta porta a traslare per l'offset della delta la forma d'onda. Se per esempio il segnale è una delta nel dominio della frequenza, e la moltiplico per delle delta a frequenza bassa, io traslo indietro in frequenza questo segnale, ottenendo, nel dominio del tempo, una sinusoida a frequenza molto più bassa:

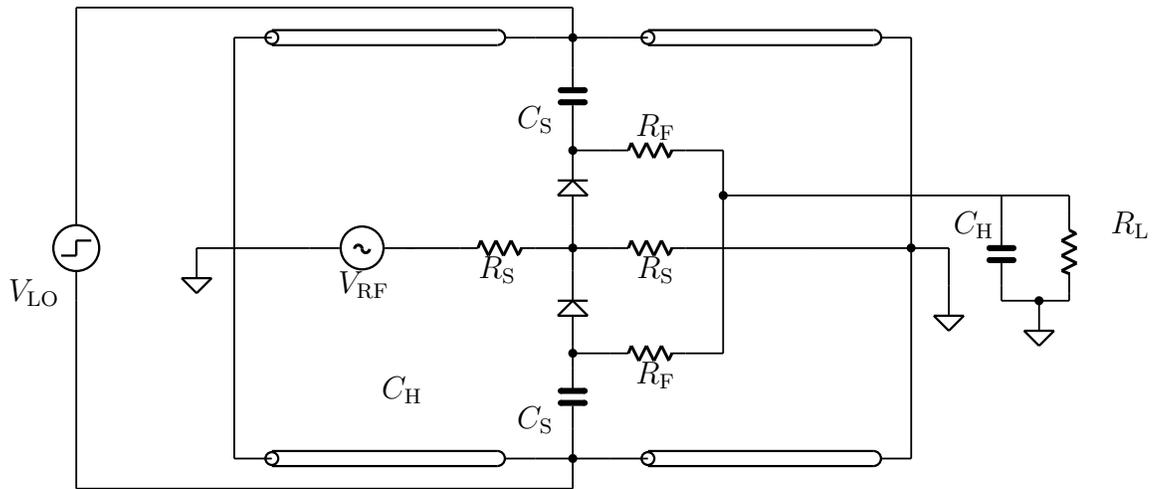


Dove sta il problema in questo fatto? Beh, finora si è parlato di campionatori ideali. Si immagni di aver a che fare con interruttori lenti, per esempio elettromeccanici: invece del treno di δ nel tempo avrò un treno di impulsi, la cui larghezza non è trascurabile rispetto alla velocità di variazione del segnale. Quello che farà un campionatore del genere è campionare in un tempo non trascurabile, dunque considerare una media di tutto ciò che entra nella finestra di campionamento (nella “porta”). In frequenza, la trasformazione di questo treno di impulsi diventerà un treno di $\text{sinc}()$, ossia di seni cardinali, cosa che comporta la formazione di un aliasing molto elevato.

Il campionatore deve essere estremamente veloce: molto, molto più veloce della variazione del segnale. Il tempo di apertura del campionatore dovrà essere trascurabile rispetto al tempo del segnale. Si consideri per esempio un segnale a 10 GHz, ossia il cui periodo è 0,1 ns: devo avere, come tempo di finestra, qualche ps; già parlare di 10 ps sarebbe troppo. Con pochi ps di apertura, si arriva a campionare anche segnali da 50 GHz (svariate decine di GHz).

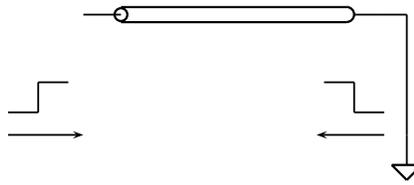
Il processo di campionamento, nel dominio della frequenza, può essere simile al mixing, al prodotto di conversione (parlando di conversione eterodina): nel caso del mixing a eterodina si mette un oscillatore locale a singola frequenza, vicina alla frequenza del segnale; nel nostro caso si hanno diverse frequenze, ossia un generatore a frequenza variabile, che farà il mixing con il segnale utile; nel caso del subsampling le frequenze possono essere anche molto basse del segnale a microonde, ma dal punto di vista spettrale capita la stessa cosa: si genera un pettine in cui la f_{sampling} vicino alla frequenza del segnale, si ha un battimento, e se ne produce una copia a bassa frequenza. Questo vuol dire che si deve garantire che l'impulso sia sufficientemente stretto, ossia che abbia un'armonica tale da finire vicino alla microonda.

Come si fa a fare un circuito con fronte di salita/discesa molto ripido e campionatore così veloce? Uno dei circuiti più utilizzati è il seguente:



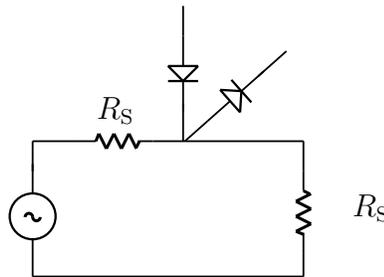
Come funziona questo circuito? Come interruttore si utilizza un diodo Schottky; se il tempo di apertura è piccolo, la quantità di carica nei condensatori dovrà essere piccola, le capacità parassite nei diodi dunque dovranno essere piccole, e dunque l'area di giunzione dovrà essere conseguentemente piccola in maniera da ridurre la capacità. Il campionatore funziona con una potenza bassa: con -10 dB_m si avrà già saturazione. Si parla di diodi che devono avere bassissima capacità parassita, poca area, ma dunque anche poca corrente, essendo l'area ridotta, e di conseguenza che reggeranno una potenza bassa. D'altro canto, essi possono essere anche utilizzati per sentire potenze bassissime: $-110 \div -120 \text{ dB}_m$. Il fatto che questo circuito sia bilanciato permette di aumentare di un poco la quantità di carica presente nei due condensatori; le due linee (da impulso a massa) sono necessarie dal momento che, avendo noi in ingresso un'onda quadra, il Γ di corto circuito

è -1; l'idea è quella di avere un impulso a onda quadra (largo) in ingresso, e di far vedere ciò:



Il fatto che il corto circuito in uscita abbia coefficiente di riflessione pari a -1 fa prendere il segnale di ingresso, lo fa ribaltare, e dunque tornare indietro molto velocemente (dipende dalla frequenza di risonanza, che dipende da λ); in questa maniera si ha in uscita a questa cosa un insieme di impulsi stretti, sovrapponendo quello in ingresso e quello in uscita: i due mezzi-impulsi (quello positivo e quello negativo) saranno quelli che polarizzeranno i diodi: ogni volta che arriva uno di questi impulsi risultanti (molto stretti), i diodi sono polarizzati direttamente, dunque conducono; questo significa che il campionario è in grado di acquisire un valore del segnale per la durata dell'impulso, dal momento che in polarizzazione diretta i due diodi si comportano come un corto circuito, la capacità dunque verrà caricata attraverso le resistenze parassite dei diodi.

Dal punto di vista della sorgente a radiofrequenza, essa è chiusa sul suo 50Ω : i diodi, normalmente, sono aperti, dal momento che senza gli impulsi essi sono polarizzati inversamente, e dunque oppongono un'alta impedenza.



Se il diodo fosse ideale, date le tensioni positive della sola radiofrequenza, esso dovrebbe condurre, per autopolarizzazione: se da un lato manca il bias, tensioni elevate della radiofrequenza fanno aumentare la tensione sui diodi, rischiando di mandarli in diretta. Quello che si fa in pratica è introdurre un bias, ossia una tensione di polarizzazione, tale per cui questa (non disegnata) sia tale da far scattare i diodi solo e solo se vi è un impulso:

Questo significa che si prende la caratteristica tensione/corrente del diodo, e, mediante il bias, la si trasla verso destra, in maniera che neanche il più

forte dei segnali a radiofrequenza sia in grado di mandare in polarizzazione diretta il diodo, dunque evitando il problema dell'autopolarizzazione. Questa cosa si ottiene regolando il bias, che regola a sua volta la soglia di conduzione, la V_γ del diodo, in modo da evitare questo tipo di avvenimento.

Cosa vuol dire ciò? Quando arriva l'impulso, il campionatore prende il segnale solo quando l'impulso supera il bias:

Fino a quando l'impulso non supera il bias non succede niente; regolando il bias dunque io sono in grado di regolare l'ampiezza dell'impulso: l'impulso non una delta esatta, ma una sorta di cono (bicchiere di champagne); regolando il bias si può aumentare o ridurre la larghezza dell'impulso preso, e dunque la banda del campionatore. Il problema che noi possiamo fare un impulso anche velocissimo, ma sarebbe meglio avere un impulso già veloce, senza andare ad "affettarlo": a parità di tempo di impulso voglio un ΔV maggiore, al fine di avere un maggior dynamic range: l'ampiezza della RF in questa maniera può essere maggiore. Questa cosa deriva dalla seguente osservazione: il fatto che il ΔV , ossia la differenza di tensione tra la soglia e il massimo dell'impulso sia grande, dà la possibilità di introdurre tensioni a RF più grandi: per superare la soglia, tanto, necessario un impulso molto più alto, dunque servirebbe una RF molto più ampia. Aumentare di molto la banda con il bias rischia dunque di portare sì a un impulso molto stretto, ma dinamicamente scarso: che aumenta di molto il rischio di mandare il circuito in autopolarizzazione. Giocare col bias possibile, ma ciò che si deve fare cercare di avere impulsi intrinsecamente stretti: giocando col bias si finisce per non avere più l'impulso, e così mandare in autopolarizzazione: l'impulso deve compensare il V_γ del diodo.

Vedremo ora come si fa a fare l'impulso più stretto possibile.

SRD: Step Recovery Diode

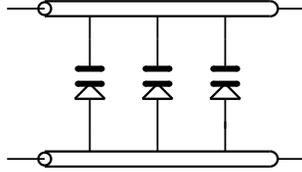
Il primo oggetto con il quale si può realizzare impulsi molto rapidi lo step-recovery diode. Esso un diodo in cui la capacità di svuotamento molto veloce; questo significa che quando mando dentro un segnale sinusoidale, dati gli effetti non lineari del diodo, si ha un trigger e il segnale esce ricchissimo di armoniche. Un diodo normale produce un segnale sinusoidale tagliato, la cui parte positiva però non esattamente sinusoidale, ma più ripida ancora; nel SRD la salita ancora più ripida, più simile all'onda quadra.

Questo SRD si può usare in questa maniera:

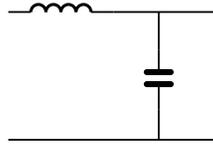
Si mette tipicamente in parallelo, in questa maniera; questo oggetto fa tagliare la forma d'onda e poi la distorce un po'. Se il segnale fatto così, il segnale in uscita ricco di armoniche. Questo non sufficiente.

Non-linear transmission lines

Quanto fatto in precedenza con l'SRD non sufficiente: necessario fare uso delle non-linear transmission lines. Cosa sono esse? Esse sono fatte in questa maniera:



Si tratta di una linea di trasmissione caricata con dei varactor. In questa maniera, si sa che la capacità di un varactor funzione della tensione; come noto, il modello di una linea di trasmissione :

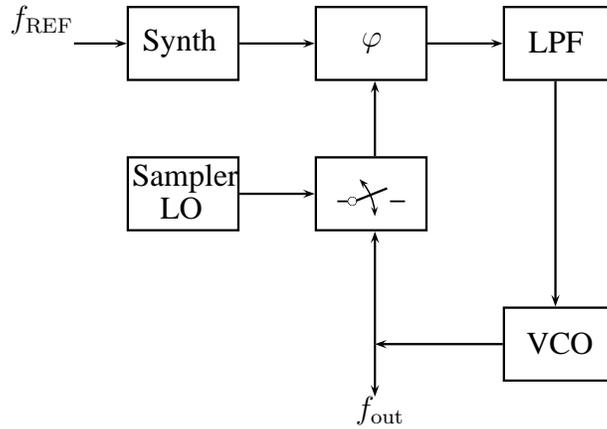


dove

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

Dal momento che la capacità variabile, la velocità di fase v_f sarà altrettanto variabile con essa. Cos' la velocità di fase? Si tratta semplicemente della velocità con la quale il segnale si sposta sulla linea, senza distorsione. Dal momento che la capacità varia con la tensione, seguendo l'andamento prima visto ($1/V^2$), al diminuire della tensione V la capacità tende ad aumentare, e viceversa; questo significa che, se $V \rightarrow 0$ (ossia se siamo sulla parte bassa del segnale sinusoidale), il segnale tende a propagarsi più lentamente, dal momento che aumenta la capacità e così si riduce la velocità di fase; dualmente con le parti alte del segnale sinusoidale, tende a propagarsi più rapidamente. Introducendo un segnale sinusoidale dunque si ha una distorsione dovuta alla differenza della velocità di fase creata dai varactor. Ciò permette di realizzare dei fronti d'onda molto più ripidi di quelli di un segnale sinusoidale.

Queste non-linear transmission lines sono state sviluppate in segreto negli anni '60: esse permettevano di avere il componente finale per la determinazione di campionatori velocissimi (migliore di quelli a SRD): si riescono a fare campionatori a qualche centinaio di GHz.



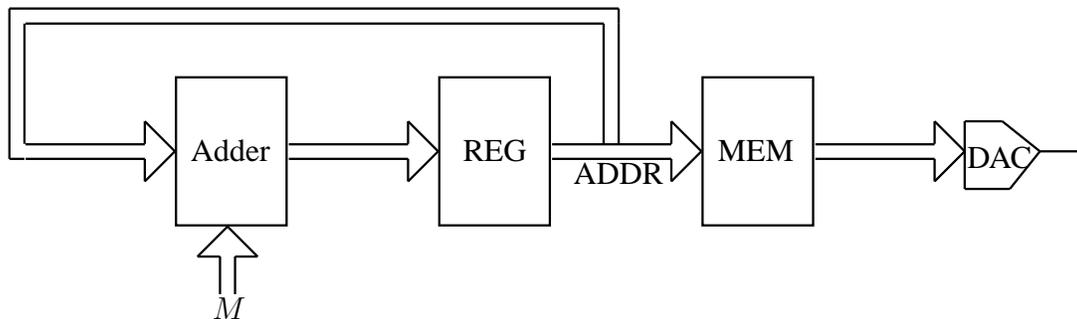
Nei generatori seri, quando si vuole avere un riferimento molto stabile, si mette un altro sintetizzatore come riferimento, a bassa frequenza: questo può essere fatto con un altro PLL a sintesi diretta, o DDS.

2.3 DDS: Direct Digital Synthesis

A questo punto si intende proporre un'introduzione al DDS, introducendo poi un insieme di schemi ad essa collegati.

Per la Direct Digital Synthesis si intende una tecnica di signal processing digitale al fine di generare un segnale con frequenza e fase regolabili, a partire da un clock fisso a una frequenza precisa. Ciò che il DDS fa è “dividere” la frequenza di clock con un certo fattore di divisione M , generalmente di 24 o 48 bit.

Il concetto su cui si basa il DDS è piuttosto elementare: si ha una memoria che contiene al proprio interno i campioni di un seno; l'uscita del sistema sarà semplicemente il seno, rappresentato in seguito alla lettura dei dati della memoria; lo schema di principio sarà il seguente:



La memoria deve essere letta introducendo al suo ingresso un indirizzo; questo è ottenuto a partire da un registro, il cui valore di uscita (ADDR) deriva da un sommatore digitale, che somma M a ogni valore in uscita dal registro; in questa maniera ad ADDR verrà sommato, per ogni colpo di clock, il valore M , e questo sarà l'indirizzo che verrà richiesto alla memoria; la memoria dunque andrà a pilotare un DAC (in realtà questo è solo uno schema di principio, ma di fatto tra memoria e DAC potrebbero esservi altri elementi, quali filtri digitali che processano il segnale), dunque, fuori dal DAC, il seno. La memoria ha, al proprio interno, 2^N celle, ciascuna delle quali contiene un campione del seno.

Se $M = 1$ e il registro alla propria uscita parte con ADDR=0, al primo colpo di clock avremo il primo byte della memoria, al secondo il secondo byte, al terzo il terzo, e così via; ciò che si può fare però è anche cambiare molto rapidamente la frequenza: la frequenza è sostanzialmente collegata con il valore di M . Se infatti prendiamo per esempio $M = 4$, significa che si legge un campione su 4 dalla memoria, e così si ha una sinusoide a frequenza diversa.

Quanto tempo ci vuole per disegnare l'intera sinusoide? Beh, si è detto che la memoria contiene 2^N campioni, che se ne prendono M a ogni colpo di clock, T_{CK} , e dunque che:

$$T_{out} = \frac{2^N}{M} T_{CK}$$

Questo è il periodo del segnale di uscita; ciò permette di avere una elevatissima risoluzione (il numero 2^N può infatti essere molto grande: si parla di ottenere risoluzione degli Hz o addirittura dei mHz); vale infatti in generale:

$$f_{out} = \frac{1}{T_{out}} = \frac{M}{2^N} f_{CK}$$

La massima frequenza di funzionamento è per $M = 2^{N-1}$; questo, sostanzialmente, per due motivazioni:

$$f_{out,max} = \frac{2^{N-1}}{2^N} f_{CK} = \frac{1}{2} f_{CK}$$

Da un lato, se si avesse una frequenza maggiore, campioneremmo solo un punto alla volta, e ciò non permetterebbe di ricostruire una sinusoide: sembrerebbe di avere una costante. A parte questo aspetto, si ricordi come sempre il teorema dei Nyquist: serve che la frequenza di clock, la frequenza di campionamento, sia almeno la doppia rispetto a quella del segnale interessato, dunque bisogna tenere conto di questo aspetto. La minima frequenza è invece quella per $M = 1$:

$$f_{\text{out,min}} = \frac{f_{\text{CK}}}{2^N}$$

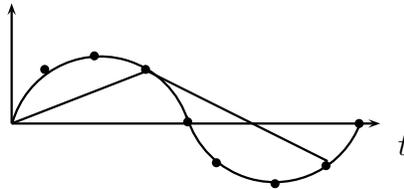
Volendo, è possibile introdurre pure un termine di fase: per fare ciò è semplicemente possibile introdurre un offset nell'indirizzo ADDR, in modo che si “parta a prendere valori” in un punto più avanzato dell'inizio.

Questa struttura presenta dei vantaggi: il sintetizzatore è molto “veloce”, nel senso che la frequenza può essere cambiata in modo pressochè istantaneo (basta cambiare M e al colpo di clock successivo si potrà già apprezzare il cambiamento). Il sistema è inoltre molto stabile, dal momento che le uniche causa di instabilità di frequenza sono da ricercarsi in f_{CK} , ossia nella frequenza di clock utilizzata:

$$\frac{\Delta f_{\text{out}}}{f_{\text{out}}} = \frac{\Delta f_{\text{CK}}}{f_{\text{CK}}}$$

Nei PLL la stabilità dipende sostanzialmente dal filtro di aggancio che intendiamo progettare: il filtro infatti (si ricordi il diagramma a farfalla) stabilisce quanto il VCO può “agganciare”; nei DDS non si ha nulla di tutto ciò, ma dunque la stabilità sarà senza dubbio molto migliore.

Ovviamente questa tecnica ha tanti vantaggi, ma anche dei problemi: questa tecnica di sintesi produce molte armoniche spurie. Come mai? Beh, si cerchi di vedere la cosa così:



Come detto questa tecnica propone semplicemente sull'uscita i campioni del seno; se si alza di molto la frequenza, se ossia si prendono pochi campioni per ciclo, ciò che “mediamente” si fa è cambiare la frequenza, ma d'altra parte si avrà una forma d'onda che assomiglierà sempre di meno a un seno; lo spettro avrà dunque molte armoniche dopo la fondamentale. La fondamentale è tanto stretta quanto quella del clock, dal momento che risente solo del phase noise aggiunto dalla parte digitale, che è bassissimo: giusto il jitter noise dei ritardi delle porte, ma dovuti al rumore, non a ritardi: la forma d'onda uscirà un minimo ritardata, ma la stabilità dipende dalle fluttuazioni del ritardo delle porte che, essendo digitali, sono poco influenzate dal rumore. Le spurie dipendono invece intrinsecamente dal funzionamento del DDS, e possono essere eliminate solo con un filtraggio, il quale viene effettuato come già accennato digitalmente, prima del DAC.

Figura 7-3

Si vedono le spurie, in seguito all'uso di un filtro passa-basso ellittico.

Si consideri ora, come esempio di sistema DDS, il seguente chip:

Figura 1-5

Questo è un sistema in cui le cose dette si ritrovano, ma con molti extra. Un esempio è il fatto che il clock è seguito da un moltiplicatore: ciò permette di ridurre la velocità del clock usato come riferimento per il sistema, cosa che potrebbe migliorarne le caratteristiche. Una volta introdotto il clock, si ha il DDS vero e proprio: di solito si tende a uscire in I/Q, ossia in modulo e fase: se si fa uscire infatti da una parte il seno, e dall'altra il coseno, si può poi proporre facilmente una modulazione di fase, e tutto ciò lavorando digitalmente; ciò richiede semplicemente l'uso di un DAC aggiuntivo per avere la doppia uscita. Prima però di arrivare ai DAC e ai moltiplicatori digitali (che verranno motivati tra breve) si hanno dei "filtri SINC inversi": dal momento che un'operazione su un segnale provoca una modifica dello spettro, ciò che si fa è equalizzare il segnale in maniera da compensare la modifica, e che dunque il segnale non risenta più della modifica; per quantizzare il segnale si usa una porta, che in frequenza va con una $SINC()$; usando questo filtro, si elimina questo effetto dallo spettro (principio della predistorsione). All'interno del blocco si ha un sommatore, che può realizzare la modulazione di fase mediante l'aggiunta di un offset. Ovviamente è possibile, come si vede, fare anche una modulazione di ampiezza, ma sempre digitalmente: si utilizza un AM modulator digitale, mediante un moltiplicatore digitale, che permette di modulare le ampiezze di I e Q, prima di mandarle ai DAC. Con tutto questo sistema, si può generare la banda base di quello che sarà un modulatore I/Q a radiofrequenza, facendo tutto digitalmente.

Questo sistema, ai tempi lo stato dell'arte, è un DDS a 300 MHz; questo significa che, se il clock è a 300 MHz, esso può funzionare al massimo a 150 MHz (per quanto si tenda sempre a fare un po' di oversampling, usando come uscita circa 1/4 della frequenza di clock); al giorno d'oggi si vedono DDS che in uscita hanno 1 GHz, dunque che lavorano con almeno 2 GHz di clock.

Indicazione, nel caso capitasse di progettare: se il sistema che si deve progettare ha una frequenza compatibile con l'uso del DDS, e il costo idem, si usi il DDS: con una discreta facilità è possibile costruire anche una radio (come detto, all'interno del blocco vi è di tutto per fare radio FM o AM); se però la frequenza in questione è sopra quella per il DDS, dobbiamo per forza ricorrere alla soluzione PLL-VCO; questa soluzione, tuttavia, molte volte è accoppiata, in diverse maniere, a una DDS.

2.3.1 Soluzioni integranti DDS e PLL

Si propongono a questo punto alcune soluzioni per accoppiare il DDS e il VCO/PLL. I DDS, come detto, hanno molti vantaggi: rapidissime variazioni di frequenza, risoluzione di frequenza, range di frequenze disponibili enorme, ecc.

DDS che lavora come VCO in configurazione PLL

Fondamentalmente si ha un PLL, nel quale si inserisce, in una posizione o in un'altra, un DDS. Si ha a che fare con qualcosa di questo tipo:

Figura 8-1

In questa figura si utilizza il DDS come divisore della frequenza del VCXO: si comporta come una sorta di divisore per M (si noti, il M del sintetizzatore, non quello del DDS). Una soluzione di questo genere, dato il fatto che la parola di tuning del DDS è molto lunga, permette di avere un divisore per M “quasi reale”. Il problema di questa tecnica sta nel fatto che questa configurazione, per quanto bella, non può essere utilizzata a microonde, dal momento che il VCO lavora, a microonde, a frequenze altissime, che dunque non sono trattabili con un DDS. Questa soluzione può essere usata a bassa frequenza, quando si vuole avere un'elevata purezza spettrale.

Soluzione con prescaler

Si propone a questo punto una variante rispetto alla soluzione precedente:

Figura 8-2

Questa soluzione è molto simile a quella di prima, ma in mezzo si ha un prescaler, ossia un divisore di frequenza: in questa maniera ciò che entra nel DDS ha frequenza relativamente bassa, ed è trattabile dal DDS. Si può, come nel caso di prima, sfruttare la proprietà dell'elevatissima risoluzione del DDS, ma in questo caso applicandola alla possibilità di gestire le alte frequenze propria del PLL.

DDS come generatore del riferimento

Un'ulteriore soluzione per l'integrazione DDS e PLL è la seguente

Figura 8-5

In questo caso si ha qualcosa di significativamente diverso da prima: ora il DDS non si utilizza come divisore di frequenza, ma come frequenza di riferimento per il PLL; si divide la RF per un numero fisso (nel caso dell'applicazione, 4096), e il reference da 14 MHz viene diviso per 64 (ottenendo un generatore da 896 MHz); dividere per 4096 significa moltiplicare la frequenza

del segnale (già divisa per 64) per 4096, ottenendo proprio 896 MHz (basti fare riferimento allo schematico di base). L'applicazione in questione utilizza il PLL a 900 MHz LMX1501A. La divisione per 64 porta a un miglioramento del phase noise di 36 dB ($20 \log_{10}(64) \sim 36$), e la moltiplicazione (risultante) per 4096 a una degradazione di 72 dB, portando a una totale degradazione, rispetto al segnale del DDS, di 36 dB.

DDS come oscillatore locale per un mixer/PLL

Si consideri infine la seguente soluzione:

Figura 8-3

Questa soluzione è particolarmente interessante: essa è un'altra possibilità che permette di accoppiare PLL e DDS, in maniera ancora diversa: essa utilizza un mixer per agganciare l'anello, downconvertendo il segnale VCO. Se al posto del mixer mettiamo un sampler, con i DDS moderni a pilotarlo (a qualche centinaio di MHz), si può usare come oscillatore locale un DDS (che sia moderno), garantendo un'elevatissima stabilità. Il sampler deve sub-campionare, e la stabilità in frequenza sarà molto più elevata di quella che si avrebbe con un semplice sintetizzatore a PLL. Questa cosa non è stata ancora applicata però negli schemi che studieremo, dal momento che un DDS a centinaia di MHz non è ancora stato applicato in questo contesto.

Lo stadio “dividi per N ” può essere uno stadio di divisione fissa oppure variabile: esso regolerebbe la frequenza di uscita del sistema in termini di “ordini di grandezza”, mentre le modifiche “fini” verrebbero fatte operando sul DDS. Le frequenze spurie del DDS vengono divise e moltiplicate nel sistema, dunque escono sostanzialmente immutate; la riduzione/aumento di spurie dipende sostanzialmente da

$$20 \log_{10} \left(\frac{f_{\text{out}}}{f_{\text{ref}}} \right)$$

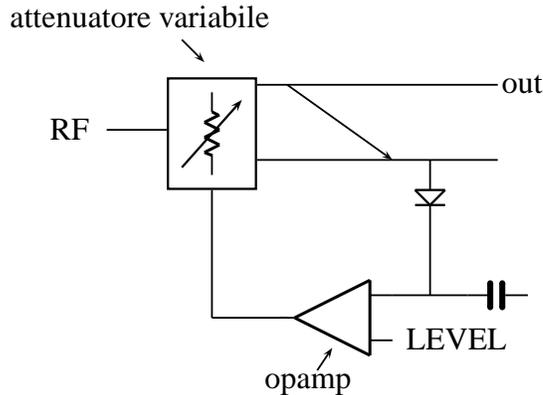
Maggiore sarà la moltiplicazione del segnale di riferimento, maggiore sarà la degradazione del segnale portato all'uscita.

Per ora, il DDS fa da oscillatore locale a un downconverter.

2.3.2 Stabilizzazione di ampiezza

Dobbiamo parlare ancora di un elemento: la stabilizzazione di ampiezza. Il generatore a microonde infatti avrà una parte dedicata alla sola sintesi del segnale e una parte che gestirà l'ampiezza del segnale, ossia un ALC: Automatic Level Control, ossia un sistema di controllo che stabilizza il segnale di uscita, in modo automatico.

In generale, un ALC si può fare in questa maniera:



La RF entra in un attenuatore controllabile, il quale è collegato a un accoppiatore direzionale: parte del segnale andrà verso l'uscita, e sarà il segnale rilevato, e parte andrà in un circuito rilevatore di picco (diodo con condensatore), il quale verrà mandato a un amplificatore operazionale: l'obiettivo sarà avere il massimo della sinusoide in uscita dall'attenuatore uguale al LEVEL (in maniera da, stabilizzando il massimo, stabilizzare anche il resto della sinusoide); l'opamp farà in modo da mantenere la tensione ai capi dei morsetti di ingresso circa nulla, in modo da porre LEVEL uguale al picco uscente dal rilevatore.

Questo sistema (magari un poco più elaborato) si usa praticamente ovunque, per la realizzazione di ALC. Il blocco critico, in questo contesto, è l'attenuatore variabile; come si può realizzare? Sostanzialmente, con un transistor in common base, del quale si cambia il punto di lavoro: in questa maniera, cambiando il bias si riesce a cambiare il guadagno del medesimo, ottenendo qualcosa in grado di mantenere costante, se controllato adeguatamente, il picco di tensione.

2.3.3 Catalogazione storica dei sintetizzatori

Prima di introdurre gli schemi pratici dei sintetizzatori, si vuole proporre una catalogazione storica.

- Fundamental single band: si tratta di un sintetizzatore che oscilla alla fondamentale in una sola banda; l'uscita è direttamente la frequenza di oscillazione del nostro sintetizzatore (ossia tutto ciò che è stato visto finora). Si usano molto spesso per esempio nei telefonini.
- Multiband fundamental: si tratta di un sintetizzatore che in uscita avrà più di una banda, e per ogni banda avrà un singolo oscillatore.

- Single band multiplier: schemi nei quali si ha a che fare con un oscillatore, seguito da un moltiplicatore; spesso si utilizzano nel campo del millimetrico.
- Multiband multiplier: quello che si usa al giorno d'oggi (soprattutto nei sintetizzatori per la strumentazione).

Oggi abbiamo a che fare con pochi sintetizzatori puramente analogici, tutti in grado di fare modulazione AM o FM, impulsata; spesso si usano i vector signal generator, che al loro interno hanno già implementato la banda base, al fine di simulare segnali di molti tipi (GSM, digitale terrestre televisivo, edge, eccetera): questo si può fare perchè sostanzialmente i frontend, l'elettronica, è uguale per tutti questi formati: cambia solo il firmware.

2.4 Analisi di schemi di generatori

2.4.1 Analisi dello schema del generatore 87MX

L'87MX è la sorgente principale di una serie di analizzatori di rete: la serie 8720. In questo caso, come si può vedere dallo schema, si hanno molti oscillatori, non seguiti da un moltiplicatore: secondo la nostra classificazione, gli 87MX sono **multiband fundamental**. Questo schema è in grado di avere, su 4 bande diverse, da 2 GHz a 20 GHz. Il segnale va attraverso due uscite: quella a sinistra (uscita +6) è quella per l'aggancio (quella che passa per il "sampler"); c'è il solito sintetizzatore a PLL, il FRAC- N , che va a finire nella shock line (che la squadra); il segnale risultante da questo processing sarà ciò che andrà a pilotare il sampler (il simbolo come quello del mixer).

I vari elementi sono su dei dies, su dei MMIC: il TC-700 è un amplificatore distribuito a MESFET (il TC-702 quasi uguale, solo un poco più potente), gli oscillatori idem sono MMIC, shockline e oscillator altri MMIC.

Sull'altra uscita (l'uscita +13) dell'oscillator IC si ha l'uscita di controllo dell'ampiezza: una serie di switch filters, un filtro generico e un attenuatore. Cosa resta fuori dall'intero 87MX ? Nell'uscita "di aggancio", il PLL; nell'uscita di controllo, rimane fuori un mixer, collegato all'attenuatore da 16 dB dopo il filtro passa basso da 5,25 GHz. Il mixer è collegato a una sorgente fissa a 3,25 GHz, e al suo interno, dall'altra parte, come si può vedere dallo schema, vi è un segnale che può avere una frequenza dai 3,25 GHz ai 5,25 GHz; questo serve perchè il mixer, utilizzando una downconversion, genera la banda più bassa. Grazie a questo trucco, si riesce a ottenere la prima banda per battimento: la banda in uscita dal generatore, dello schema, in questo modo, parte da qualche decina/centinaio di kHz. Nel caso in cui si voglia

generare una frequenza bassa, dunque, si collega il solid state switch +15 al ramo superiore (quello con il filtro a 5.25 GHz), con il generatore $2 \div 5$, 25 GHz collegato in ingresso; si farà la downconversion e in uscita il secondo switch +15 sarà collegato.

I filtri sono tutti filtri passa-basso: essi servono per eliminare le armoniche spurie, le seconde armoniche, prima di collegare all'uscita: l'oscillatore ha infatti una fondamentale (data dallo YIG), ma si han anche seconde armoniche.

Il resistore variabile è fatto a transistor, per modificare l'anello di ALC (controllo dell'ampiezza di uscita).

Ultima osservazione, il 27 kHz, che modula l'interruttore: questo generatore serve a creare lo stesso effetto del modulatore a 1 kHz in laboratorio, sulla linea a fessura: permette di migliorare il funzionamento del ALC, "accendendo e spegnendo" in continuazione, modulando dunque in modo impulsato (l'interruttore o è acceso o è spento), rendendo il segnale più alternato, più semplice da trattare dal loop del ALC.

2.4.2 Analisi dello schema del generatore Anritsu MG369XA MM

A questo punto analizziamo lo schema del frontend del MG369. Per analizzarlo, considereremo le varie parti, pezzo per pezzo.

- Si consideri (2-7) lo schema a blocchi generale del sistema: si possono identificare, da sinistra verso destra, si ha un microprocessor, la parte di interfaccia I/O, la parte di power supply (fondamentale, in tutti questi schematici).
- Si consideri (2-27) il frontend del sintetizzatore: questa è la parte che si occupa della gestione dello YIG e della produzione dell'uscita.
 - Il blocco A9 è lo YIG oscillator, ossia è un monolitico realizzato all'incirca come è stato descritto precedentemente; lo schema non fornisce dettagli ulteriori sul monolitico; avrà 2 oscillatori, in modo da andare con uno da 2 a 8,4 GHz, col secondo da 8,4 a 20 GHz.
 - A10 è un primo switched filter assembly: un insieme di filtri che si sceglie (l'uso dei PIN switches come in questo caso è tipico della Anritsu). Si va ad A12 verso il sampler, e si viene dal ALC, ossia dal level control (blocco A6): gli switch sono controllati dall'anello di controllo del livello, e dunque A10 esce su un filtro passa basso a 20 GHz, atto a tagliare le armoniche superiori (dopo i filtri, che

già di loro hanno provveduto a tagliare le armoniche superiori del segnale che si vuole sintetizzare).

- A11, ossia la option 5, permette di ottenere una down-conversion che può essere utilizzata per ricavare la banda base in uscita dal generatore. Questa si ottiene (da 0,01 GHz a 2 GHz) facendo battere il segnale con un generatore locale a 6,5 GHz fisso (ottenuto su un riferimento a 500 MHz passante dal blocco A3). Esiste una soluzione alternativa per realizzare l'alta frequenza: usare l'option 4 invece che l'option 5. Questa è basata su un DDC (Digital Down Converter), che permette di fare una divisione, in digitale, del segnale, e filtrandolo (si ha una rete di banchi di filtri per il tal proposito). Il livello di controllo della banda base si fa con un accoppiatore direzionale in banda base, realizzato mediante elementi a parametri concentrati (induttanze e condensatori), con un diodo per detectare da 10 MHz a 2 GHz.
 - Proseguendo sulle alte frequenze, mediante in A10 si può selezionare o la banda base o la banda; dopo di esso ancora il fatidico passa basso di cui già si è parlato, poi lo switched doubler module: un sistema che, a scelta, o moltiplica per 2 la frequenza (40 GHz), oppure non fa nulla, in A12.
 - Si ha un ulteriore cammino possibile, opzionale: si ha la possibilità di introdurre una banda base che va da 0,1 Hz a 2 GHz, mediante un DDS, in option 22 (blocco A8), gestito da una coppia di diplexer. Il DDS che produce da 0,1 Hz a 10 MHz è pilotato dall'uscita a 26,84 di A3.
 - Ulteriore opzione è lo step attenuator: attenuatore a celle elettromeccaniche, a larghissima banda (deve funzionare fino a 40 GHz), che serve per generare un segnale di ampiezza anche molto ridotta: prima si stabilizza a 10 dB_m, dopo dunque si può attenuare con questo blocco. Può essere utile per esempio per simulare il path di antenna, situazione nella quale si deve arrivare anche con -100 o -120 dB_m. Di solito ha 90 dB di attenuazione possibile.
- Variante dello schema è quello che usa il source quadrupler, che con un segnale a 16,8 GHz fa una moltiplicazione di frequenza per 4, ottenendo da 0,1 Hz a 65 GHz.
 - Si consideri a questo punto la parte di aggancio di questo sistema, ossia quella in cui il segnale viene agganciato (2-11).

- Si ha (A9) lo YIG assembly, che contiene l'oscillatore a YIG e il relativo PCB; lo YIG ha tre segnali di pilotaggio: tune, main, bias. I main coil drivers pilotano il coil, mentre il fine tune driver è ciò che viene fuori dall'aggancio di fase.
- A7 YIG lock: l'aggancio viene fatto attraverso un sampler, il quale viene pilotato da un SRD (che in realtà è una shockline), con un segnale a bassa frequenza (fino a 245 MHz). Ciò crea armoniche da 1,9755 GHz a 20 GHz, per generare una IF che va da 21,5 a 40 MHz. Nel phase detector si fa entrare una IF alla stessa frequenza, di riferimento, e il fatto che questo è un anello mantiene ciò che è in uscita dal mixer costante; il segnale in uscita dal phase detector sarà un segnale di errore, un segnale di correzione per ciò che si ha da mettere nello YIG (e verrà introdotto attraverso il fine coil driver, in maniera da cambiare di pochissimo il campo magnetico dello YIG), e dunque in questo modo si fa l'aggancio. Solitamente lo YIG ha due ingressi di controllo, per i coil, ma nel caso ne avesse solo uno si può lo stesso fare il controllo fine, introducendo un sommatore di segnale di fine tuning all'unico ingresso.
- Il segnale a 245 MHz e quello a 40 MHz di riferimento devono essere generati in qualche maniera; al sampler non si può mandare un segnale da un mixer, perchè non ha senso: vorrebbe dire che si ha un altro segnale a microonda, che per a sua volta non potrebbe esser stato degenerato da nulla. I reference e il pilotaggio del sampler cambiano: per il sampler cambia perchè bisogna centrare in maniera corretta l'armonica, ma la differenza tra la n -esima armonica e il segnale potrebbe cadere in un punto sbagliato, e campionare correttamente a segnale fisso potrebbe richiedere eccessiva risoluzione. In altre parole, permette un fine tuning sull'armonica prodotta: visto che le armoniche si ottengono per battimento, nel caso l'altra non fosse molto precisa, si può variare la frequenza dell'altro fattore, in maniera da agganciare in maniera corretta (dobbiamo essere in grado di cambiare la IF in maniera da prendere l'armonica giusta).
- A4 (il Coarse loop) è un altro generatore a PLL, a 490 MHz, diviso 2: come mai si divide per 2, e si fa così? Beh, questo perchè in questo modo si riduce un pochino il jitter, ossia il phase noise del sistema, garantendo contemporaneamente una possibilità di spaziatura delle armoniche molto accurata. La divisione nel PLL del Coarse Loop si fa con un mixer, pilotato da un segnale a frequenza fissa (500 MHz, ricavato da un 100 MHz). Il frequency

divider in A4 potrebbe essere fatto a DDS, tanto le frequenze in gioco sono basse; in un sintetizzatore moderno.

- La parte veramente “diversa” nei sintetizzatori moderni è l’uso del DDS: a partire da un clock di 300 MHz (blocco A3), si usa un DDS, adesso, per fare la sintesi dei segnali interessati. A3 è il blocco che effettua la sintesi di frequenza per A4, A7, A9 di questo schema. In ingresso si ha un segnale di controllo della CPU al fine di gestire la sintesi di frequenza (che tendenzialmente come detto è a DDS), quindi in uscita si avranno i segnali alle frequenze dette; a seconda di cosa si vuole produrre in uscita, si può fare il conto.
- Intuitivamente, come funziona tutto questo sistema? In uscita si può avere un segnale dai 2 ai 20 GHz; questo può essere generato all’inizio dallo YIG, che si suppone vicino ad esso; lo YIG è tunato da un segnale tale per cui si abbia una certa armonica in uscita dal SRD (shockline); scelta l’armonica più vicina, si vede qual è il segnale di ingresso al SRD che porti ad avere la tal armonica più vicina al numero che vogliamo; questa andrà moltiplicata per 2, e sarà la frequenza che entrerà nel mixer di A4; questo si andrà a downconvertire con 500 MHz, e il frequency divider avrà un M tale da riportare a 1 MHz (segnale di riferimento) il suddetto valore di frequenza in uscita (IF) dal mixer.
- Tutti i sintetizzatori moderni hanno un VCXO (Voltage Control Cristal Oscillator), spesso stabilizzato in temperatura, con possibilità di aggancio a riferimento esterno (che potrebbe essere un cesio, un rubidio). Al giorno d’oggi, con la diffusione del segnale GPS, si ha a disposizione in modo molto semplice un segnale di tempo accurato a 10^{-12} s. Vi sono dei generatori, dei clock locali, che sono agganciati automaticamente al segnale GPS per potersi sincronizzare. Quando si vogliono fare cose professionali si usano oscillatori al rubidio (con stabilità elevatissime ma per poco tempo) che grazie all’aggancio GPS vengono agganciate ai cesi degli orologi atomici della costellazione GPS, allineati tra loro: si ha la possibilità di creare un riferimento molto stabile in frequenza grazie al GPS, che si può usare per entrare (di solito non si usa un generatore a microonde attaccato al GPS). Quello che si fa di solito è usare REF IN e REF OUT: si può usare, con la stessa time base, più oscillatori tra loro: collegando i riferimenti tra loro, si usa la stessa base tempi, lo stesso riferimento temporale per tutti i generatori, semplificando di molto la vita.

- La parte relativa alle modulazioni FM e PM si ottiene sommando il segnale modulante esterno nel FM control path: invece che pilotare con il solo PLL lo YIG si pilota anche con un segnale esterno FM.
- La parte relativa alle modulazioni AM si gestisce a livello di ALC, ossia all'interno del controllo automatico di ampiezza, sommando all'esterno e all'interno del blocco i segnali. Questi segnali di solito si introducono nel rear panel del nostro sintetizzatore.
- Si può anche avere un generatore di impulsi per avere la modulazione impulsata. Questa può essere usata, per esempio, per sincronizzare le sweepate di uno schermo a raggio catodici, CRT.

Si introduce a questo punto un insieme di proprietà importanti per determinare le prestazioni del sistema.

- Single sideband phase noise test (tabella 3-2): questa tabella permette, a partire dell'offset dalla carrier, di quantificare le prestazioni di phase noise;
- Per le performances di velocità, si faccia riferimento a Phase-locked step sweep mode: si hanno specifiche su come capire le caratteristiche.
- Si parla di stabilità: Internal time base stability : 10^{-9} ; questa è quella di un quarzo molto buono; si noti che se la risoluzione è 10^{-12} , si hanno comunque meno cifre significative di quelle effettivamente visualizzabili.
- Switching Time: il tempo che si impiega per cambiare a frequenza dei sintetizzatori; per passare da una certa frequenza a un'altra, per esempio, non si impiega più di 40 ms. Questa è una specifica critica: essa è quella sostanzialmente che fa da collo di bottiglia per gli analizzatori di rete: essi devono essere utilizzati in concomitanza con un generatore di segnali, dunque si dovranno fare misure solo dopo che la frequenza si sia stabilizzata, dunque non prima che passi lo switching time. Ciò che limita tutto ciò è lo YIG: ipotizzando (numeri tipici) 80 mH e 8 Ω , in risultante si hanno

$$f = \frac{1}{2\pi} \frac{L}{R} \sim 15\text{kHz}$$

Ciò che limita il nostro sistema è il controllo magnetico di corrente dello YIG.

2.4.3 Analisi dello schema del generatore Agilent E8257N

Si consideri una seconda tipologia di generatore di segnale, studiando un altro schema. Questo schema è di Agilent, molto simile al precedente, se non per alcune peculiarità che andremo ad analizzare. Si ha a che fare con un oscillatore basato a YIG, il quale però arriva solo a 10 GHz (a differenza di Anritsu che può arrivare anche oltre): per fare 20 GHz servirà già un moltiplicatore. Vediamo i dettagli della macchina.

- Foglio 1:
 - Il blocco A5 costituisce il sistema di sampling e di Coarse tune, ossia l'anello per la gestione dello YIG oscillator; il sampler, realizzato mediante un mixer, è indicato (in basso a destra) con una S. Appena sopra si può anche vedere il phase comparator, nello stesso blocco. La IF viene sintetizzata come nell'altro schema (si ha un mixer con un segnale da 500 - 1000 MHz che può arrivare con due modalità: uno per divisione (modo A), e uno a partire da un segnale da 1 GHz (il segnale 1 GHz Output to A5 sampler proveniente dal blocco A7 reference) che fa 1 GHz diviso 4 che va a mescolarsi con 1 GHz, ottenendo un 750 MHz; questo 750 MHz genererà l'oscillatore locale LO per abbassare il segnale del sampler (ricopre sostanzialmente il compito dei 500 MHz di prima).
 - Blocco A6: qua si vede il blocco che genera il segnale di riferimento principale del sistema; in questo blocco si può verificare la prima grande differenza con il sistema Anritsu: anziché mediante DDS, in questo caso il segnale principale viene realizzato mediante un anello a VCO. Il segnale in uscita dal VCO va a entrare in un prescaler e verrà a formare un anello con il quale si realizzerà la sintesi del segnale.
 - A9: nel blocco A9 vi è qualche dettaglio in più, rispetto allo schematico Anritsu, sul metodo di pilotaggio dei coil dello YIG (lo YIG oscillator è nel blocco A28): si vede che il sistema è pilotato in corrente dal segnale "TUNE", che entra nel sommatore dal basso (il generatore di corrente è il quadrato con la freccia, collegato al sommatore), più un segnale di correzione (quello che entra nel sommatore dall'alto) che arriva da sopra, al fine di poter avere il "fine tuner". Si ha la possibilità di modulare FM, mediante il generatore collegato alla "FM Coil", con all'ingresso dell'FM il segnale che parte dal phase detector.

- Fogli 2, 3, 2+3 (far partire il foglio 3 con la fine del foglio 2 di fianco, in modo da vedere intero A10 - ALC):
 - Blocco A35: 3-20 GHz I/Q modulator; si tratta di un blocco che realizza la demodulazione I/Q, a partire da un segnale, in uscita dal 20 GHz doubler A29 (blocco abbastanza simile, almeno concettualmente, al doubler Anritsu; l'implementazione è diversa), produce la componente in fase e la componente in quadratura. Il sistema Anritsu non aveva l'I/Q modulator. In ingresso dunque si ha, in A35, il segnale che verrà demodolato I/Q; a parte ciò, è anche possibile che questo non venga suddiviso nelle componenti, ma esca (uscita J1) direttamente al blocco A30; qui una parte del segnale, dopo un filtro di modulazione, degli accoppiatori, e un attenuatore, andrà alla RF Out; l'altra parte invece entrerà nell'anello ALC di controllo automatico dell'ampiezza, passando in un attenuatore controllato in tensione.
 - Si ha una trattazione particolare della banda base, da 0 a 3 GHz: il blocco A14 genera direttamente la banda base. In queste macchine, dette PSG Vector RF path (vector signal generators), si ha anche una parte di generazione di base-band con tutto un insieme di controlli (dei quali non ci interessiamo), dove il segnale viene preso o dall'esterno, o dal generatore interno, e viene mandato all'I/Q modulator interno.
 - Il blocco A13 è il I/Q multiplexer: prende il segnale I/Q, lo fa passare in un sommatore, ci somma alcuni ritardi (in maniera da compensare i ritardi interni del sistema), aggiunge tempi di sweep nel caso di sweepate e così via. Dopo tutto questo giro nel blocco A13 il segnale esce (3 to 20 GHz) e va a inserirsi nel A35 (I/Q modulator).
 - Si noti, nel blocco A8, la presenza di un secondo I/Q modulator: esso, come si può notare dallo schema, prende il segnale di ingresso dalla banda base. Questo è un I/Q modulator dedicato alla sola banda compresa tra i 0 e i 3 GHz.

Capitolo 3

Analizzatori di rete

3.1 Introduzione agli analizzatori di rete

Ciò che si sta per introdurre è uno strumento atto a caratterizzare il segnale in un circuito; la caratterizzazione del segnale può essere di due tipi: lineare (dato un segnale a un tono in ingresso al circuito, avere un solo tono in uscita al medesimo), o a più toni (un solo tono di ingresso, diverse frequenze in uscita). Al fine di effettuare una caratterizzazione lineare, è necessario utilizzare l'analizzatore di reti.

Come noto, ogni circuito lineare può essere rappresentato mediante una rappresentazione Thèvenin, sia nel caso di circuito a una sola porta, sia nel caso abbia svariate porte; vale sempre la rappresentazione matriciale

$$\underline{V} = \underline{Z}\underline{I} + \underline{V}_0$$

Caratterizzare un circuito con \underline{Z} , \underline{S} , è sostanzialmente molto simile: i risultati sono tra loro tutti collegati, mediante trasformazioni bilineari di un qualche tipo. Nel nostro caso, ciò che contraddistingue i vari parametri, le varie rappresentazioni tra loro, è collegato alla loro "utilità" rispetto al modo in cui li misuriamo: ciascuna rappresentazione è definita a partire da un certo numero di operazioni da effettuare per misurarle, e dunque a seconda di come si effettuano le misure si hanno diverse utilità nell'usare l'una o l'altra relazione. Nel caso di misure a bassa frequenza, per esempio, la matrice \underline{Z} va benissimo: si annulla la corrente sulla porta di uscita, si eccita la porta di ingresso, e così si misura la tensione "su circuito aperto".

A seconda della misura da ottenere (impedenze, parametri scattering), sarà necessario imporre diverse condizioni di sorgente e di carico; si misureranno, per un singolo tono, tutti i parametri alle varie porte, e quindi poi si

cambia la frequenza, ripetendo per ciascuna frequenza nel range interessato le misure dei parametri delle varie porte.

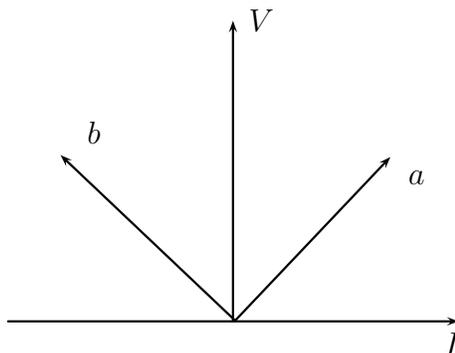
Quando si parla di radiofrequenza, si hanno diverse questioni: serviranno di sicuro misure vettoriali, dal momento che si devono misurare grandezze in modulo e fase; quando si parla di frequenze elevati, come le radiofrequenze, la struttura con cui si ha a che fare è distribuita: si hanno tensioni e correnti che dipendono sostanzialmente dalla posizione nella quale ci si trova. Non ha senso misurare tensioni o correnti: quelle che si misurerebbero sarebbero le quantità alle porte, ma pochi millimetri dopo si misurerebbe qualcosa di completamente diverso, vanificando il senso della misura. Tensioni e correnti infatti si definiscono a partire dalle soluzioni dell'equazione d'onda, le quali sono costituite da funzioni ondulatorie al variare della posizione spaziale. Parlare invece delle "soluzioni dirette" dell'equazione d'onda, ossia dei modi progressivo e regressivo che vengono a propagarsi nella struttura, è più appropriato per la nostra situazione: l'unico termine che ci deve preoccupare è un po' di sfasamento.

Questo ci porta a capire che la scelta naturale per fare misure a radiofrequenza è quella dei parametri scattering; dividendo tensioni progressive e regressive per la radice di una certa resistenza di riferimento, si applica sostanzialmente che si ha qualcosa del tipo:

$$V^+(z) = \frac{V(z) + Z_r I(z)}{2}$$

$$V^-(z) = \frac{V(z) - Z_r I(z)}{2}$$

Questo, come si può vedere, è un banale cambio di base dato da una trasformazione lineare delle già note $V(z)$ e $I(z)$: una semplice "rotazione" del piano $I(V)$.



Si immagini, su questo piano, di avere una resistenza: una caratteristica lineare. Nel nuovo dominio, come noto, si avrà ancora una retta, che però ora sarà detta “coefficiente di riflessione”:

$$b = \Gamma a$$

Si consideri una linea di trasmissione: volendo calcolarne i parametri scattering, la matrice risultante è:

$$\begin{cases} 0 \\ e^{-jkz} \end{cases} e^{-jkz} 0$$

Una matrice antidiagonale, ma questo è vero se e solo se la linea è caricata sulle impedenze di riferimento della matrice scattering, e che la Z_r sia uguale alla Z_∞ della linea. Questo è un punto critico: non è assolutamente scontato che ciò sia vero, e che dunque la linea introduca un semplice sfasamento, senza avere della riflessione. Non è tutto: si è dimostrato che ciò che si riesce a misurare non è precisamente il coefficiente scattering nel senso “classico”, quello introdotto da Kurokawa negli anni '60, bensì qualcosa di diverso (se ne discuterà ancora tra breve).

Definizione sempre più importante è quella dei parametri S differenziali; come mai? Beh, al giorno d'oggi, sia che si parli di telecomunicazioni sia che si parli di computing, si parla di microonde; HDMI, per esempio, ha delle velocità di trasmissione dei GiB/s, e le interfacce dunque lavoreranno sostanzialmente a microonde. Nei sistemi con i quali al giorno d'oggi si ha a che fare, si hanno molti pin e una frequenza elevata, le piste “parlano” tra loro, ossia tendono ad accoppiarsi; nasce dunque il problema della **signal integrity**, ossia il segnale viene distorto dal crosstalk o dai vari altri fenomeni, quali riflessioni multiple.

Si ha a che fare sostanzialmente con alcune problematiche.

- Come è possibile generare i segnali e campionare?
- Dov'è il piano di riferimento del sistema? (Ossia, come si può calibrare il sistema?);
- Qual è l'impedenza di riferimento del sistema?
- Quali sono le proprietà di questi parametri e qual è il loro legame con la propagazione ondosa? La risposta a questa domanda è abbastanza semplice, per chi conosce la teoria delle linee di trasmissione: i parametri differenziali si comportano con le stesse caratteristiche di propagazione

di a e b , ma questo sono nel caso in cui le impedenze di riferimento siano:

$$Z_{c,jk} = \frac{R}{2}, \quad \text{per il modo comune}$$

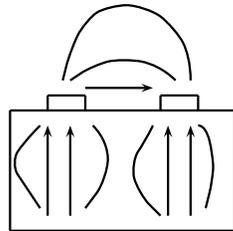
$$Z_{d,jk} = 2R, \quad \text{per il modo differenziale}$$

- Come si può fare in un circuito nel quale alcune porte sono differenziali e alcune porte single ended? (Esempio pratico l'amplificatore operazionale: due pin costituiscono una porta differenziale, mentre in uscita si ha un singolo pin, un'uscita single-ended.)

Si ha a che fare anche con problematiche secondarie.

- Non stiamo utilizzando substrati a microonde: stiamo usando FR4, che è un substrato scarso, non adatto a ciò che si intende fare.
- Sono fatti con interconnessioni non idonee alle microonde.

Come mai si utilizzano solo modi differenziali, nelle strutture in questione? Si consideri qualcosa di questo tipo:



Il modo differenziale è quello che fa parlare le due piste; il modo comune è quello secondo cui ciascuna delle due piste parla con la massa. Il modo differenziale è l'unico nel quale si può introdurre dell'informazione dal momento che il substrato più in commercio è FR4: esso ha perdite ed è anisotropo, dunque introdurre in esso informazione è un suicidio; nessuno però al mondo ha la potenza commerciale per togliere dal commercio FR4; se si riuscisse a fare un substrato come l'allumina con il prezzo dell'FR4, si potrebbe sostituire e quindi risolvere questo problema.

Per questo motivo si parla di parametri differenziali; si utilizza un cambio di base del tipo:

$$a_{d,jk} = \frac{1}{\sqrt{2}} (a_j - a_k)$$

$$b_{d,jk} = \frac{1}{\sqrt{2}} (b_j - b_k)$$

$$a_{c,jk} = \frac{1}{\sqrt{2}} (a_j + a_k)$$

$$b_{c,jk} = \frac{1}{\sqrt{2}} (b_j + b_k)$$

Una nota: queste formule valgono se e solo se ha a che fare con quattro pin, quattro porte: si fa in modo da accoppiare le porte tra loro a due a due, e finita lì; se si ha a che fare con più porte, la faccenda si complica terribilmente. Ciò che si deve fare è definire un insieme di coppie differenziali e di coppie di modo comune: dato un sistema con N porte, se ne possono avere P differenziali e $N-P$ di modo comune, single-ended (come l'operazionale). Per gestire sistemi misti (sia differenziali sia single-ended) è necessario utilizzare una matrice scattering $\underline{\underline{S}}^0$ detta **matrice scattering mixed-mode**; essa è collegata alla matrice $\underline{\underline{S}}$ classica secondo una relazione del tipo:

$$\underline{\underline{S}}^0 = \underline{\underline{M}} \underline{\underline{S}} \underline{\underline{M}}^{-1}$$

dove $\underline{\underline{M}}$ è una opportuna matrice di cambio di base dalla matrice scattering tradizionale a quella mixed-mode.

Purtroppo, c'è qualcosa in più rispetto a ciò: chi propose questo cambio di base (Eisenstadt), non tenne conto di un fatto molto importante, ossia del fatto che le caratteristiche di propagazione dei modi differenziali e dei modi comuni non seguono esattamente quelle delle a e delle b , se non in un caso specifico: quello per cui valgono le relazioni già dette:

$$\begin{cases} Z_{c,jk} = \frac{R}{2} \\ Z_{d,jk} = 2R \end{cases}$$

solo in questo caso la matrice mixed-mode ha senso, per come è definibile. Partendo a questo punto dal seguente set di equazioni:

$$a_{d,jk} = \frac{V_{d,jk} + I_{d,jk} Z_{d,jk}}{2\sqrt{R_{d,jk}}}$$

$$b_{d,jk} = \frac{V_{d,jk} - I_{d,jk} Z_{d,jk}}{2\sqrt{R_{d,jk}}}$$

$$a_{c,jk} = \frac{V_{c,jk} + I_{c,jk} Z_{c,jk}}{2\sqrt{R_{d,jk}}}$$

$$a_{c,jk} = \frac{V_{c,jk} - I_{c,jk} Z_{c,jk}}{2\sqrt{R_{d,jk}}}$$

La risultante è una matrice $\underline{\underline{S}}^0$ che sarà a blocchi:

$$\underline{\underline{S}}^0 = \begin{bmatrix} \underline{\underline{S}}^{dd} & \underline{\underline{S}}^{cd} \\ \underline{\underline{S}}^{dc} & \underline{\underline{S}}^{cc} \end{bmatrix}$$

Questa matrice è costituita da quattro blocchi di matrici. Questo è ciò che è stato proposto nell'articolo originale che ha definito i parametri scattering differenziali. Questo formalismo ha alcuni problemi, tra cui uno evidente: ora esso è perfetto per la rappresentazione di sistemi differenziali, ma non è un vero **mixed-mode**, nel senso che si tratta solo di una descrizione per sistemi differenziali, che non tiene conto del fatto che un sistema può avere, come già detto, sia porte differenziali sia porte single-ended; per ciascuna porta differenziale sarà necessario definire due parametri, mentre per quelle single-ended uno solo. Una matrice che tenga conto di tutto dovrà avere una forma del genere:

$$\underline{\dot{b}} = \underline{\dot{S}} \underline{\dot{a}}$$

dove

$$\underline{\dot{a}} = \begin{bmatrix} \underline{a}_d \\ \underline{a}_c \\ \underline{a}_s \end{bmatrix}$$

Dove \underline{a}_d è il vettore delle onde progressive nelle porte differenziali, contribuiti di modo differenziale; \underline{a}_c il vettore delle onde progressive nelle porte differenziali, contribuiti di modo comune; \underline{a}_s il vettore delle onde progressive nelle porte single-ended. Allo stesso modo, si può parlare dei vari vettori delle onde uscenti:

$$\underline{\dot{b}} = \begin{bmatrix} \underline{b}_d \\ \underline{b}_c \\ \underline{b}_s \end{bmatrix}$$

Questa rappresentazione dei parametri scattering, in grado di tenere conto sia delle porte differenziali sia delle porte di modo comune, è detta **generalized mixed mode**, è una matrice costituita da 9 blocchi di sottomatrici, dunque con qualcosa del tipo:

$$\begin{bmatrix} \underline{\underline{S}}^{dd} & \underline{\underline{S}}^{dc} & \underline{\underline{S}}^{ds} \\ \underline{\underline{S}}^{cd} & \underline{\underline{S}}^{cc} & \underline{\underline{S}}^{cs} \\ \underline{\underline{S}}^{sd} & \underline{\underline{S}}^{sc} & \underline{\underline{S}}^{ss} \end{bmatrix}$$

quindi

$$\begin{bmatrix} \underline{b}_d \\ \underline{b}_c \\ \underline{b}_s \end{bmatrix} = \begin{bmatrix} \underline{\underline{S}}^{dd} & \underline{\underline{S}}^{dc} & \underline{\underline{S}}^{ds} \\ \underline{\underline{S}}^{cd} & \underline{\underline{S}}^{cc} & \underline{\underline{S}}^{cs} \\ \underline{\underline{S}}^{sd} & \underline{\underline{S}}^{sc} & \underline{\underline{S}}^{ss} \end{bmatrix} \begin{bmatrix} \underline{a}_d \\ \underline{a}_c \\ \underline{a}_s \end{bmatrix}$$

Anche questa mixed mode generalizzata può essere ricavata a partire dalla matrice $\underline{\underline{S}}$ standard, però attraverso una trasformazione più complicata:

$$\underline{\underline{S}} = \left[\underline{\underline{\Xi}}_{21} + \underline{\underline{\Xi}}_{22} \underline{\underline{S}} \right] \left[\underline{\underline{\Xi}}_{11} + \underline{\underline{\Xi}}_{12} \underline{\underline{S}} \right]^{-1}$$

Dove $\underline{\underline{\Xi}}_{21}$ è una matrice idonea in grado di tenere conto sia della trasformazione delle porte differenziali sia di quelle single-ended.

I problemi non sono ancora finiti: anche una volta giunti a questa generalized mixed mode matrix, in realtà c'è un problema, ossia il set di equazioni dal quale siamo partiti. Ai tempi della definizione dei parametri scattering, purtroppo non si è tenuto conto di come effettivamente si potesse misurare la matrice $\underline{\underline{S}}$ e i suoi componenti, e di un fatto: quello che effettivamente si misura non sono le onde di potenza nel senso noto, ma ciò:

$$a_{d,jk} = \sqrt{R_{d,jk}} \frac{V_{d,jk} + I_{d,jk} Z_{d,jk}}{2 |Z_{d,jk}|}$$

$$b_{d,jk} = \sqrt{R_{d,jk}} \frac{V_{d,jk} - I_{d,jk} Z_{d,jk}}{2 |Z_{d,jk}|}$$

$$a_{c,jk} = \sqrt{R_{c,jk}} \frac{V_{c,jk} + I_{c,jk} Z_{c,jk}}{2 |Z_{c,jk}|}$$

$$a_{c,jk} = \sqrt{R_{c,jk}} \frac{V_{c,jk} - I_{c,jk} Z_{c,jk}}{2 |Z_{c,jk}|}$$

Questo set di equazioni tiene conto della fase dell'impedenza caratteristica, dal momento che si distinguono Z e R , dove quest'ultima è definita come la parte reale di Z . Nel caso in cui il substrato abbia perdite (caso realistico, come si è detto prima, dal momento che si ha a che fare con il pessimo FR4), le relazioni prima ricavate non vanno bene, dal momento che non si ha la semplificazione con il denominatore di \sqrt{R} , e così le equazioni ricavate sono inesatte. Una teoria ricavata a partire da queste equazioni è di sicuro molto più vicina a quanto si misura effettivamente.

Riprendiamo a questo punto come ultimo concetto il procedimento di misura dei parametri scattering; si parta dalla seguente formula:

$$S_{ij} = \left. \frac{b_i}{a_j} \right|_{a_i, i \neq j = 0}$$

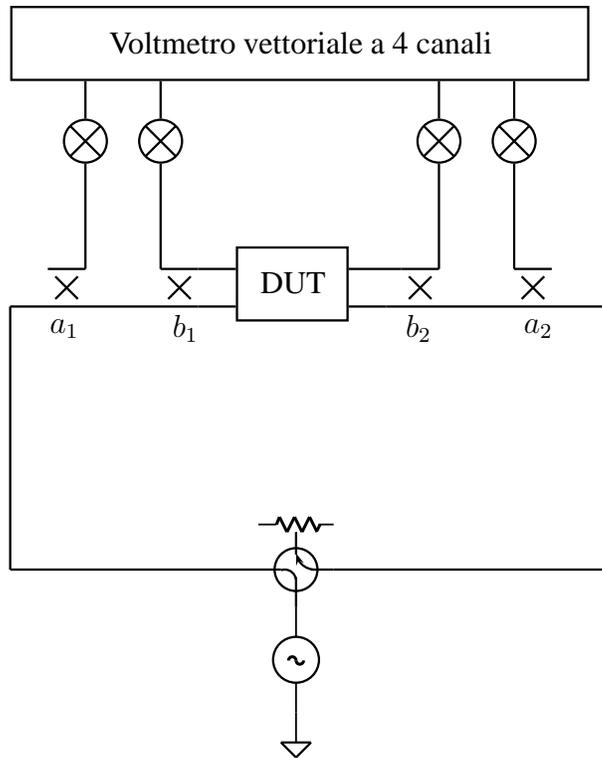
Come si riassume questo procedimento? Prima di tutto si carica ogni porta del DUT (Device Under Test) sulla sua impedenza di riferimento tranne la i -esima; si eccita dunque la porta j -esima, e si vede cosa esce alla i -esima.

3.1.1 Schemi di base dei VNA (Vector Network Analyzers)

Dato questo insieme di nozioni, si cerchi di ricavare alcuni schemi di base del VNA.

Schema 4-sampler

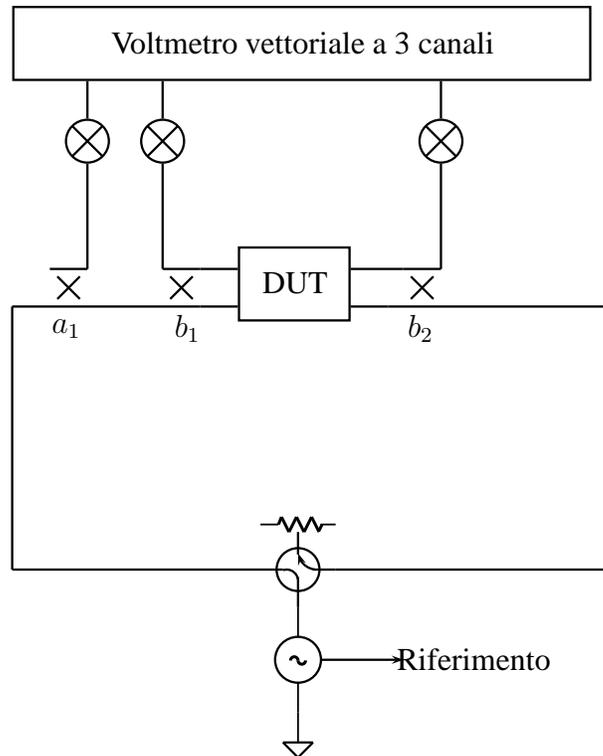
Consideriamo a questo punto lo schema fondamentale del VNA, quello probabilmente intuitivamente più semplice. Si parte da un DUT, ossia da un Device Under Test, ossia dal blocco che intendiamo caratterizzare secondo i parametri scattering; come si può leggere dalla definizione dei parametri scattering, è necessario eccitarlo a una qualche porta; questo porterà, di sicuro, ad avere una certa sorgente a radiofrequenza. D'altro canto, la porta non eccitata (le porte) dovrà essere chiusa sulla sua impedenza di riferimento; usando uno switch si fa in modo sia da avere, con la stessa configurazione, qualcosa di cambiabile in maniera molto versatile: a seconda della posizione dello switch si può scegliere di caricare (caso 2-porte) la porta 1 al generatore e la porta 2 all'impedenza di riferimento, o viceversa. Terminato ciò, si devono misurare i segnali **incidenti** e **riflessi** alle due porte: al fine di poter rilevare questi segnali senza "disturbare" il circuito, è necessario introdurre dei dispositivi in grado di prelevare i segnali medesimi: quattro **accoppiatori direzionali**, i quali effettuano la **signal separation**. Gli accoppiatori permettono di misurare i segnali, che poi verranno mandati in quattro mixer, i quali fanno da **microwave receivers**: prendono i segnali a microonda, li tirano giù di frequenza, e permettono di mandarli in un misuratore a bassa frequenza, quale un voltmetro vettoriale a quattro canali.



Il voltmetro vettoriale ovviamente lavorerà alla IF dei mixer.

Schema 3-sampler

Un'osservazione rispetto alla misura precedente: quando si intendono misurare i parametri scattering di un DUT, in realtà non siamo interessati alla misura di tutte le a : non siamo interessati alla a_i , ossia all'onda progressiva della porta che intendiamo misurare (quella che non si chiude nel carico), poichè siamo solo interessati alla sua onda regressiva, b_i . Questo ci dice che il procedimento precedente effettua più operazioni di quelle che in effetti ci interessano. Vale dunque anche uno schema di questo genere:



Questo schema ha un accoppiatore direzionale e un mixer (in totale 3, da qui il nome 3-sampler) in meno: si ha un solo riferimento e due altri mixer.

Purtroppo, questo schema, come si vedrà, ha molti svantaggi. Il vantaggio sostanziale sta nel fatto che dunque il segnale di riferimento per il voltmetro vettoriale si può prendere, invece che mediante un quarto accoppiatore direzionale, direttamente dalla sorgente a radiofrequenza, a monte degli switch anzichè a valle.

Le sorgenti di solito sono a frequenze altissime, sintetizzate, con combinazione DDS+PLL; sono a larga banda, abbastanza livellate, **non** hanno problemi di phase noise, dal momento che di solito la frequenza che andiamo a eccitare è isolata rispetto a ciò che misuriamo, dunque anche se la frequenza non è stabilissima, balla un pochettino, il risultato va benissimo. Deve però essere altamente ripetibile.

3.2 Introduzione agli schemi fondamentali

Si consideri a questo punto di essere oltre alle due porte:

Si consideri come esempio la seguente soluzione (Agilent PNA-X):

Vi sono due ricevitori per ogni porta, e in questo caso due sorgenti, ognuna per una coppia di porte. Si potrebbe arrivare a fare una sorgente per ogni porta, soluzione Schwartz/Ballman.

3.2.1 Schema Anritsu 12 porte 65 GHz

Sono sufficienti quattro porte? No. Per andare oltre alle quattro porte, si arriva a fare strutture molto complicate. Uno degli analizzatori di rete più sofisticati presenti attualmente in commercio è il seguente Anritsu 12 porte 65 GHz:

Questo sistema, a 12 porte, è sostanzialmente basato su di una switching network. Questa struttura è abbastanza simile (concettualmente) alla struttura 3-sampler. Nonostante questo fatto, si ha una certa simmetria tra i due insiemi di porte: c'è una linea di simmetria tra le due parti del sistema. Il sistema esce con 12 porte, e per ogni porta si ha un accoppiatore direzionale, come nella situazione 3-sampler (uno solo, non due); l'accoppiatore direzionale viene multiplexato, nel senso che lo switch permette scegliere una delle due uscite. Ognuno di questi moduli ha due porte, e contiene al suo interno uno switch atto a fare questo multiplexing dei segnali in ricezione. Al livello superiore, vi è un altro switch, il cui obiettivo è quello di portare la sorgente in una direzione o in un'altra. Ciascun modulo (quelli al livello più basso, in ciascuno dei quali vi sono due porte) ha due cavi: uno che porta il segnale, e l'altro che preleva il segnale. Si manda tutto ciò ad una switching unit più grossa, che ha due switch per parte: ancora una volta uno per la sorgente, uno per la ricezione, questa volta però switch 3 to 1: due per la parte a sinistra, e due per la parte a destra della linea di simmetria. Passati da qui, la ricezione va a finire su un sampler dell'analizzatore di reti, e la potenza viene distribuita dall'altro switch (per ciascuna parte). Si hanno due segnali di riferimento: uno per la parte a sinistra, e uno per la parte a destra, dal momento che questo analizzatore è ricavato a partire da un analizzatore completo a 2 porte, 4-sampler:

Questa situazione non è nè quella con due accoppiatori per porta nè quella con un solo riferimento: si hanno due riferimenti, uno a destra e uno a sinistra, ma d'altra parte non si hanno, per ciascuna porta due accoppiatori, ma solo uno. Questa cosa, vedremo, porta ad avere semplificazioni al momento della calibrazione dello strumento.

3.2.2 Schema Agilent

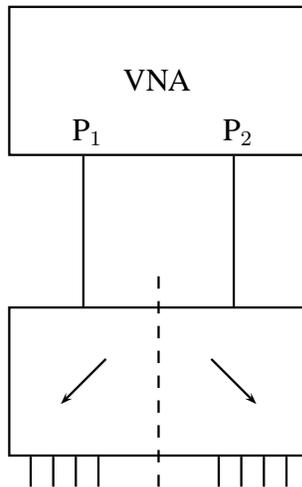
Anche nel caso dello schema Agilent si ha qualcosa di abbastanza simile:

Nel caso della configurazione Agilent si hanno 10 porte invece che 12, anche in questo caso con una netta suddivisione tra sinistra e destra; la particolarità della soluzione sta nel fatto che si sfruttano anche due porte nell'analizzatore di reti, cosa che non si fa nell'Anritsu: volendo sfruttare le porte in alto bisogna portare fuori il segnale dell'analizzatore di reti, accedendovi direttamente, e ciò nell'Anritsu non è stato fatto. Si hanno come prima due riferimenti, due ricevitori, switchati da una switching network.

3.2.3 Ipotetico schema

Riassumendo: come già detto servono una sorgente, due test-set, due microwave downconverters e un voltmetro vettoriale. Volendo poi passare da due porte a molte porte, quello che si fa nella maggior parte dei casi è utilizzare una switching network (o, a bassa frequenza, semplici multiplexing, che ad alte frequenze sarebbero proibitivi).

Si propone a questo punto il seguente schema:



Mettere cioè le due porte del VNA, tanti switch e collegare direttamente gli switch alle porte. Come mai non si può fare ciò? Perché questa cosa non funziona: se si avessero solo due switch 4 to 1 non sarebbe possibile misurare due porte dalla stessa parte; a parte ciò, c'è una faccenda più complicata: il vero problema è che tarare in maniera semplice uno strumento realizzato con una rete di questo tipo non è possibile. Se quindi da un lato questo strumento costa meno di quelli precedentemente introdotti, dal momento che non hanno gli accoppiatori direzionali, calibrarlo è molto difficile. Questi oggetti vengono utilizzati ma solo a bassa frequenza, per esempio quando si devono analizzare le prestazioni delle reti di distribuzioni di segnali quali

quello televisivo nei condomini (che tanto sono 1 to N). L'accuratezza di questi analizzatori, però, è molto scarsa.

La cosa migliore è avere almeno due sampler per porta, ma si possono ottenere ottimi risultati anche con le soluzioni a simmetria. Le soluzioni basate su switching network saranno sicuramente molto più lente. Si torna dunque al discorso della sorgente: è necessario, per avere analizzatori di rete veloci, sorgenti veloci.

3.3 Analisi di alcuni schemi di network analyzers

Analizzeremo, in questa sezione, sostanzialmente due modelli (facendo cenno a un terzo).

Si consideri a questo punto una prima introduzione allo schema iniziale del 8720 (pag 12-3): si possono identificare la sorgente sintetizzata, la parte di separazione del segnale, il ricevitore (la parte a mixer), dunque il display e le restanti interfacce con l'esterno. Si noti però una peculiarità, utilizzata molto frequentemente negli strumenti moderni: si noti che "phase lock", ossia l'anello di aggancio per il sintetizzatore, viene preso dal ricevitore. Questa è una cosa che si fa in tutti i sintetizzatori moderni, e ha un'importantissima conseguenza: nel caso in cui il DUT generi un segnale, questo tipo di configurazione non funziona. Se il DUT genera una frequenza, l'unico modo di sincronizzare questa frequenza generata, questo segnale, con quello del sistema, richiederebbe un modo di agganciare le due basi tempi, i due segnali; se il DUT non ha possibilità di agganciare la base tempi, si otterrebbe qualcosa di analogo all'aprire l'anello di aggancio, perdendo la stabilizzazione. L'unico modo sarebbe collegare in qualche modo le due basi tempi, ma ciò non è sempre possibile. Questa cosa porta anche vantaggi: non si usa più infatti il sampler all'interno del sintetizzatore, ma si usa quello già presente nel ricevitore.

Queste note sono state tratte dalla service guide del 8720; in realtà, queste considerazioni valgono in egual misura anche per il PNA.

3.3.1 Analisi dello schema del PNA

A questo punto analizziamo, punto per punto, lo schema a blocchi del PNA, facendo riferimento sia a uno schema a blocchi semplificato che a uno un po' più elaborato.

- La sorgente di segnale è, come sempre, basata su uno YIG; l'oscillatore a YIG è contenuto nel blocco A12. Si può vedere che si ha una sorgente doppia: si hanno due sorgenti, ossia quella dell'oscillatore locale, e quella dello YIG. Questo schema sarà mixer-based, essendo i punti critici ad alta frequenza (come si può vedere da Foglio 1). Lo YIG utilizzato ha una frequenza che va dai 3 ai 10 GHz (riferimento a YTO). Il PNA riesce ad andare fino a $60 \div 70$ GHz, ottenuta moltiplicando i $3 \div 10$ GHz dello YIG. PMYO è un oscillatore che si usa per downconvertire fino a ottenere la banda base per il segnale; questo è un oscillatore fisso, probabilmente realizzato mediante dielettrico. Sul foglio 1 si ha una tabella che permette di identificare le varie bande presenti nel sistema. Dopo la parte di source 10, si hanno due moltiplicatori, sempre interni al blocco A12: è possibile moltiplicare fino a 8 volte il segnale, arrivando fino a 80 GHz teoricamente (ma, vedendo la tabella, in realtà il massimo è 70 GHz). Si parla di "SOMA: Source Oscillator Multiplier Assembly": questo è un insieme che realizza un blocco sorgente e moltiplicatore.
- Blocco A17: LOMA (Local Oscillator Multiplier Assembly): in ingresso il LOMA ha A8, il FRAC-N, nel quale, per sintetizzare la frequenza, si utilizza un VCO per fare da oscillatore locale. In uscita si ha un ulteriore moltiplicatore per 2, A18, dunque in A19 uno splitter, e in A20 il LODA: Local Oscillator Distribution Assembly; questo si può fare con un power splitter, che "splitterà" la potenza dell'oscillatore locale. Si moltiplica anche qui per 8, ma qua si parte da 3 GHz: $3 \times 8 = 24$. La frequenza del LO arriva al massimo a 24 GHz (da tabella a 23,33 GHz), quella della sorgente a 70 GHz; come mai? Questo perchè si utilizza un mixer in subharmonic: si usa il subharmonic mixing, ossia il mixer non si usa alla fondamentale, ma con una subharmonic. Lo svantaggio di questa cosa, aldilà del phase noise, è il fatto che usare in subarmonica un mixer porta ad avere un minore dynamic range. Dal momento che l'alternativa sarebbe stata realizzare anche il LO a YIG, soluzione estremamente costosa, questa va ancora bene. Al giorno d'oggi comunque si trovano VCO anche a frequenze elevatissime, dunque non è un problema.
- Il blocco A10 realizza il riferimento di frequenza, con un OCXO o in alternativa con un riferimento esterno, frequenza di riferimento di 10 MHz. In alcuni strumenti, lo switch è comandato dal fatto che si senta o meno un fronte dall'esterno: se si ha un clock dall'ingresso, lo switch lo sente e automaticamente si imposta su di esso. Questa cosa c'è di sicuro nel 8720 che analizzeremo in seguito, forse anche nel PNA in analisi. Il

riferimento di 10 MHz, da ovunque esso venga, si aggancia mediante il solito anello ad aggancio di fase con un riferimento a 100 MHz, ottenuto con divisori per 2 e 5 (come si può vedere dall'anello interno ad A10). Fuori da questo anello, a partire da questo 100 MHz, mediante due divisori selezionabili mediante switch si possono ricavare dei riferimenti di frequenza: uno da 8,333 MHz che verrà mandato all'anello di gestione dello YIG (A11), uno da 4 MHz (che poi va a diventare da 20 MHz), che servirà per realizzare la frequenza del secondo oscillatore locale: su Foglio 3, dove vi è un maggiore dettaglio, si vede che questa uscita è diretta verso un "2nd LO \times 4": questa è la frequenza utilizzata per realizzare la seconda downconversione; $\times 4$ è scritto dal momento che questo tipo di schema si basa sullo schema concettuale a 4 mixer, dunque ci saranno quattro fili che andranno a fare da LO per quattro mixer.

- Cos'è A35, dal quale viene il segnale che viene sommato? Studiando lo schema a pagina 5-7, possiamo intuire che A35 sia la board dei receiver.
- Blocchi da A27 a A30 e vicini: si può vedere che si hanno 4 mixer, che sono quelli che si occupano della prima downconversione. Si hanno quattro uscite (A, R₁, R₂, B), che vanno a un multiplexer (dentro A35), che sceglie uno dei quattro segnali e su questo si potrà fare il phase lock, agganciandosi dunque a esso (come si era detto prima, questo è quello che significa "agganciarsi sul ricevitore").
- Blocco A20 (riferimento a foglio 2): come si può vedere, nel blocco di distribuzione del local oscillator si hanno sei amplificatori: se una volta di solito si usava solo un power splitter, in realtà si fa anche condizionamento del segnale. Stessa cosa prima dei mixer: anche qui si hanno degli amplificatori aggiuntivi. Tutti questi amplificatori, prima dei mixer, devono essere amplificatori a basso rumore e banda larghissima, dal momento che sono posti nel path del segnale: questi possono avere frequenze fino a 70 GHz, dunque dovranno essere amplificatori distribuiti. Il fatto che gli amplificatori siano low-noise, porta vantaggi sotto il punto di vista del dynamic range.
- Il blocco A49 è l'IF multiplexer: è quello che esegue l'operazione di multiplexing che arriverà sulla A35: receiver motherboard.
- Per quanto riguarda la seconda conversione, si ha qualcosa come lo schema 5-22: l'oscillatore locale è collegato a uno dei quattro mixer di ricezione tramite A49; la seconda conversione sdoppia in due rami

la conversione. Il segnale in ingresso, diviso tra due mixer, subisce il seguente processing: in un ramo la IF non viene sfasata, nel secondo sì, di $\pi/2$; questo può essere interpretato in due maniere:

- questo schema è lo schema Hartley IRR: se pensato come un ricevitore, essendo questa una seconda conversione, questo schema è utile al fine di performare le reiezione delle frequenze immagine;
- questo schema è in grado, con questo trucco, di trovare le componenti in fase e in quadratura del segnale; queste vengono, alla fine, sommate, in maniera da misurarle in seguito con il blocco A6-SPAM: in un analizzatore di reti il nostro obiettivo è quello di misurare I e Q: quello che si fa è separarle, amplificarle separatamente, e quindi misurarle. Quello che si fa di fatto con questo schema però è mandare tutto in convertitori A/D.

Una nota: solitamente, le componenti I e Q sono in banda base, in DC; in questo caso non lo sono, e sono a 41 kHz, prima di entrare nel ADC. C'è anche un altro aspetto: trattando il segnale in I/Q, il segnale si amplifica “meglio”. La misura I/Q si fa digitalmente e non analogicamente, dal momento che si fa nel blocco A6, di signal processing. Dal momento che la cosa è digitale, modulo e fase si possono misurare con moltiplicatori digitali. La misura si fa con un NCO (Numerical Control Oscillator, una sorta di DDS a frequenza fissa), utilizzato come oscillatore locale; prendendo il seno e il coseno, moltiplicandoli digitalmente, si può realizzare questa operazione, tirando fuori parte reale e immaginaria del segnale.

Tutto ciò viene fuori in modo molto accurato per un motivo semplice: con questo sistema noi sappiamo già, a priori, dove stanno le frequenze. Se non sapessimo infatti dove sono le frequenze, teoricamente dovremmo fare la FFT e trovare i picchi, in modulo e fase, di dove stanno; questo purtroppo non è di solito così facile, dal momento che fare la stima spettrale non è un'operazione banale: usando il metodo “scolastico” di solito non si trovano mai i risultati corretti. Fare la FFT implicherebbe avere una quantizzazione “implicita”, ossia dovremmo avere il segnale quantizzato, al fine di applicare numericamente l'algoritmo FFT. Fare il prodotto per seni e coseni di una frequenza già nota, invece, permette di proiettare il segnale di ingresso su elementi già noti di una base di frequenze, non dovendo più stimare dove la proiezione è avvenuta: lo sappiamo già. I fasori ottenuti con questo procedimento sono assolutamente stabili.

- Si noti, in A6 (il modulo di signal processing) il IF calibration signal: questo è un segnale molto accurato che permette di compensare gli errori di linearità e di non-linearità propri del ADC, e del signal processing successivo. Si noti ciò: ogni tanto l'analizzatore di reti, dal momento che si potrebbero avere segnali in ingresso con dynamic range molto elevato, potrebbe variare automaticamente, mediante un'opzione software che controlla delle parti hardware, il fattore di scala. Questa cosa è da disabilitare, nel caso in cui si intenda misurare delle potenze invece che dei rapporti: periodicamente infatti si avrebbe questo cambio di scala, rendendo la misura insignificante.
- Pagina 5-16: questo è il test set, al quale sono collegate le due porte; si ha uno switch, l'attenuatore, il bias T, e l'accoppiatore direzionale; un solo accoppiatore direzionale: essendo questo lo schema a 4 mixer, l'accoppiatore nel blocco A25 prende solo il segnale "riflesso", su entrambe le porte. I segnali incidenti si prendono attraverso un power splitter: Nel 8510 il power splitter era realizzato mediante un power splitter; ora si fa con dei monolitici che riescono a fare questa operazione. Si noti che non tutti i test set sono a power splitter: in quelli più vecchi vi erano solo accoppiatori direzionali.
- Pagina 5-19, figura 5-4: questa è un'opzione interessante, per cui si hanno dei contatti che ponticellano il frontale. Se si deve comprare un analizzatore, lo si compri con queste opzioni: se si ha necessità di usare l'analizzatore con opzioni particolari, per esempio nelle quali bisogna sostituire i coupler, avere dei multiport, andare più su di frequenza, questi ponticelli sono utilissimi.

3.3.2 Analisi dello schema del 8720

Si analizza a questo punto un secondo schema, sotto alcuni punti di vista differente dal precedente.

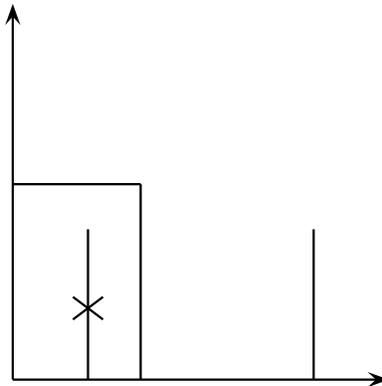
- Pagina 12-13: come si può vedere in figura 12-3, si ha il reference a VCXO, il source YIG oscillator, e il R-sampler: questa è una sampler-based machine e non una mixer-base machine; questo ci suggerisce il fatto che l'oscillatore locale di riferimento, in questo caso, sarà a bassa frequenza (come si può vedere nel FRAC-N, infatti, si ha un VCO-out da 60 a 240 MHz).
- Pagina 12-17, figure 12-4: si hanno due oscillatori a YIG: A54 e A55. A55 è esattamente identico a quello che avevamo studiato precedente-

mente; lo YIG2 è più moderno, e va da 20 a 40 GHz. La banda base si fa sempre per battimenti con un oscillatore fisso (A53 con A57).

- Pagina 12-22: signal separation. Questo schema solitamente non viene utilizzato con l'opzione 400 (pag. 12-20), che è quella con quattro canali; tipicamente non si ha neanche il ET model, ma si ha uno switch dopo lo step attenuator (A74 transfer switch, per passare dalla porta 1 alla porta 2), con un solo riferimento.
- Pagina 12-24: un ricevitore, in questo caso a 3-sampler, a doppia conversione. Si hanno tre sampler sia in prima sia in seconda conversione, IF digitale, e un solo ADC. La seconda conversione tira giù tutto a 4 kHz, la prima a 10 MHz. Questo strumento è stato il primo a fare usare la IF digitale, ossia il primo a fare signal processing digitale (calcolare le componenti I e Q); come già detto, aveva un solo A/D converter.
- Foglio 4 (sono cose già scritte ma in dettaglio): si ha il blocco di sorgente (SOURCE), e l'oscillatore locale è a 240 MHz, che entra in uno step recovery diode (A14, A13, A52).
- Lo YIG (A55) ha al solito da 2,55 a 20 GHz, e può andare a battere con la banda base, phase lock (A11) con reference da un lato e 10 MHz in ingresso, 1 MHz dal reference (A12).
- Foglio 5: si ha il blocco di receiver, nel quale si ha il R_1 loop standard, che va al sampler di ricezione (A64, R_1 sampler), e i due sampler di segnale (A e B); passando dalla signal separator, si può vedere che in uscita essa va a A e B. Per fare la signal separation ci sono due options: 007 o 085. Esse sono basate sulla sostanziale differenza nello switch: nel caso della option 007 si ha un interruttore solid-state, con dei bias-tee per il disaccoppiamento delle componenti DC e segnale; nel caso della 085 uno switch meccanico. Le options 012 sono i ponticelli di cui si parlava prima: direct sampler access.
- Foglio 6: in questo foglio si ha uno schema diverso, ossia l'8722; questo è uno schema in cui ci sono delle varianti rispetto al precedente. In questo schema si hanno ancora due YIG, il A25 fa da divisore di banda, ma in questo caso si hanno 4 sampler: questo schema ha l'opzione 400 (foglio 7). Si hanno 4 sampler, ma solo 3 downconverter in seconda battuta, sulla seconda conversione (non è un 4+4, ma un 4+3); questo dunque porterà problemi. Come si può vedere dai sampler R_2 e R_1 , essi vengono commutati, selezionati, o sul A4 o sul A5 (i quali sono downconverter di seconda battuta). A4 e A5 vengono usati per la ricezione, mentre

A6 viene utilizzato per fare ricezione e generazione del riferimento per il phase lock: A6 è il sampler del reference (come si vede nel blocco A10, digital IF).

- Si faccia riferimento a foglio 6 e pagina 5-3 (quest'ultima più semplice): quello che noi facciamo in pratica è misurare introducendo un segnale nel DUT, cosa ne esce e cosa riflette. Seguendo i percorsi dei segnali, si può intuire che R_1 misura l'onda di potenza a_1 ; A misura quello che torna indietro dalla porta sinistra del DUT, dunque b_1 ; dalla porta a destra del DUT, esce b_2 , la quale va nella porta B. Ciò che posso fare è misurare A, B, R_1 o R_2 muovendo gli switch S11 S12 S21 S22, e quindi, sfruttando A4, A5, TRL CAL, fare a_1/a_2 . Utilizzando gli switch si può scegliere di misurare tutti i rapporti.
- Foglio 8: power supply. Si vuole introdurre l'idea dietro il power supply. Come si può vedere a sinistra, l'alimentatore è basato su un regolatore switching (parte A15 di preregolazione): questo è un regolatore di tipo mezzo-ponte: due switch, due condensatori + 1. In fondo, però, si hanno anche dei regolatori lineari. Come mai tutto ciò? Quello che si fa è avere uno switching per gestire le tensioni di ingresso, il 50 Hz, e dopo un regolatore di tensione. Questo si fa dal momento che, sotto il punto di vista armonico, un regolatore di tensione lineare è sostanzialmente equivalente a un filtro, che taglia la componente a 50 Hz per lasciare solo una componente continua. Un convertitore switching prende il segnale (DC + 50 Hz) ed effettua un campionamento (una frequenza di campionamento tipica potrebbe essere 100 kHz); questo significa che, sostanzialmente, si ha un'armonica a 0 Hz (la DC), e la successiva sarà traslata a circa 100 kHz, eliminando il 50 Hz. Il regolatore lineare, agente come detto da filtro, avrà una reiezione molto più efficace rispetto al caso 50 Hz, dal momento che deve filtrare un'armonica a frequenza molto più elevata rispetto a prima.



Alimentatori di questo tipo dunque vengono fatti al fine di ottenere, in uscita, tensioni continue pulitissime, realizzate mediante la combinazione delle due tecniche di raddrizzamento dell'onda.

3.4 Taratura degli analizzatori di rete

In questa sezione, che come si vedrà sarà molto corposa, verranno presentate, dopo un insieme sottocapitoli introduttivi al problema, le principali tecniche di taratura applicabili sugli strumenti in studio: gli analizzatori di rete.

3.4.1 Problematiche introduttive

Digital signal processing

Si parla di digital signal processing sostanzialmente dal 1985, quando uscì il modello 8510, modello di analizzatore di reti un po' "consunto", ma che sicuramente può rappresentare le problematiche. L'idea sostanzialmente è quella di permettere l'elaborazione dai dati acquisiti dallo strumento fuori da un convertitore A/D. Si ha uno schema a blocchi di questo tipo:

In uscita dal ADC si ha una memoria che contiene dati RAW, ossia i dati in uscita dallo strumento. Questi dati sono insignificanti, dal momento che qui vale il discorso del salumiere: di fatto le misure, senza che lo strumento sia stato calibrato, sono completamente insensate. Ciò che bisogna fare è prendere i dati, e correggerli mediante il blocco ERC, ossia il blocco **error correction**, che utilizzerà come due ingressi per l'appunto i dati RAW, e quindi i coefficienti di correzione dell'errore (CAL coefficients), contenuti anch'essi in una memoria. I dati ottenuti da questa elaborazione sono i dati corretti, che dunque possono essere mandati a un blocco di trasformazione, un tempo realizzato mediante un processore matematico, che realizzava una specie di trasformata di Fourier. Fuori dal blocco di trasformazione vi sono due strade: una porta verso una memoria, una verso un formatting block: esso è necessario dal momento che come dati abbiamo sostanzialmente un array di numeri complessi, di fasori (poichè trasformati nel dominio di Fourier); formattarli significa renderli visualizzabili su carta di Smith, carta polare, e così via. Si fa dunque lo scheduling e lo smoothing dei dati, per poterli visualizzare: tutto ciò è processing locale.

Tutto ciò al giorno d'oggi si va via software, mediante per esempio C++; questo schema a blocchi tuttavia è molto importante dal momento che anche oggi i software di elaborazione sono basati, concettualmente, su di esso: può essere necessario avere accesso ai dati RAW o a quelli corretti, per fare

diversi tipi di analisi o dimostrazione; può essere per esempio utile cambiare tecnica di taratura, e per tarare è comunque necessario accedere ai dati RAW in qualche maniera: molte volte si è costretti ad adattare l'hardware alle tecniche di taratura disponibili, dunque le misure potrebbero non essere ottimali, e in certi problemi la cosa potrebbe essere molto importante. Ciò è fondamentale dal momento che i risultati utili si ottengono a partire da un de-embedding (error correction) dei dati RAW, "filtrati" attraverso un modello di errore.

A parte il voler cambiare la taratura, di solito la maggior parte delle volte è fondamentale accedere ai dati corretti: corrected data.

Tutta questa introduzione serve per comprendere una cosa assolutamente fondamentale: il processo di taratura è assolutamente necessario, al fine di dare un senso alla misura dei parametri scattering: se non si facesse la taratura, la misura, dal momento che la "tara" non è conosciuta, sarebbe così errata da non poter neanche comprendere l'ordine di grandezza ottenuto.

Problema dell'interfacciamento

Un ultimo problema da conoscere prima di parlare seriamente di taratura è il cosiddetto problema dell'interfacciamento: molte volte, quando si intende caratterizzare un blocco, non abbiamo la possibilità di accedere direttamente ai piani di riferimento del dispositivo; volendo per esempio caratterizzare un connettore, può essere che si debbano costruire attorno ad esso delle interfacce di test, delle **fixture**. Il problema di questa cosa è il fatto che esse non saranno direttamente attaccate al coassiale, e dunque il piano di riferimento dovrà essere portato il più possibile vicino al dispositivo.

Un modo per risolvere il problema potrebbe essere quello di misurare i parametri scattering dell'intera strumentazione e "sottrarre", nel senso di de-embeddare, con solo la differenza tra riferimento e uscita, la sola fixture. Ci sono, per poter misurare dal riferimento all'uscita della fixture, alcuni metodi: uno può essere quello di cercare di mettere gli "standard" dove servono, ricavando i dati necessari mediante simulazioni FEM del pezzo esterno, o uso di una board che ha tutto attaccato. Sul wafer di silicio si ha qualcosa di simile: si hanno dei collegamenti esterni.

3.4.2 Introduzione alla taratura

Riassumendo ciò che è stato detto finora, si ha a che fare con un certo insieme di problematiche.

1. Il primo problema riguarda il fatto che gli strumenti, quali gli accoppiatori direzionali, che non hanno direttività infinita e comportamento

variabile in frequenza. Questo significa, dal salumiere, che “la bilancia ha dei problemi”: la strumentazione non è ideale.

2. Gli switch non sono ideali, anch'essi frequency dependent; d'altra parte essi sono estremamente ripetibili, dunque almeno qualche vantaggio lo hanno.
3. Abbiamo problemi dei reference plane: essi dipendono sostanzialmente dalle fixture che introduciamo.
4. Vi sono problemi di second'ordine, che però non sono molto importanti, nonostante un profano cercherebbe di riparare essi: phase noise e altri. Questi problemi sono di second'ordine dal momento che se si sbaglia a inserire i coefficienti di taratura (vedremo dopo cosa sono essi), la misura non avrà senso, non sarà solo un po' sbagliata; dal salumiere, questo significa sbagliare a calcolare la tara, non capendo più niente di quello che pesiamo.

Quando si parla di misure mediante analizzatori di rete, le cause di incertezze, di errore, sono:

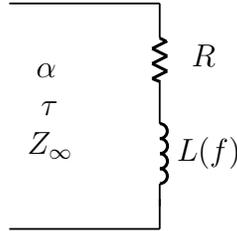
- al 90% da errori sistematici, ossia da errori di taratura; questi, in quanto sistematici, sono errori stabili nel tempo, che possiamo eliminare in modo efficace con una buona taratura;
- al 10% (o meno) sono gli errori random: rumore, cattiva cura del laboratorio (si usino, per questi, alcuni accorgimenti, quali usare sempre la chiave dinamometrica per stringere i coassiali, usare connettori compatibili tra loro, pulirli, tenere accese le strumentazioni, quali sorgenti e analizzatori, al fine di non avere effetti negativi causati dal drift del quarzo).

3.4.3 Standard di calibrazione

In tutte le tecniche di taratura usate al giorno d'oggi, si utilizzano sempre gli stessi standard. Questi sono sostanzialmente i seguenti: load, open, short, thru, reflect. Questi sono sostanzialmente gli standard che possono essere collegati su di una porta. Verranno a questo punto descritti uno a uno, al fine di capire cosa essi rappresentino.

Standard short

Per short si intende un modello di corto circuito; al fine di definire un corto circuito sono necessari 5 parametri: tre relativi alla linea di trasmissione che si sta usando, uno relativo all'impedenza di carico:



Gli standard da specificare sono τ , ossia il ritardo introdotto dalla linea (strettamente collegato con la lunghezza della linea: maggiore è τ , maggiore sarà la lunghezza della linea); Z_∞ (impedenza caratteristica della linea); α (parametro di perdita, di loss). Per quanto riguarda l'impedenza di carico, si ha una resistenza R e un'induttanza L che in verità è una funzione della frequenza f del segnale; essa di solito si identifica nella seguente maniera:

$$L(f) \sim L_0 + L_1 f + L_2 f^2 + L_3 f^3$$

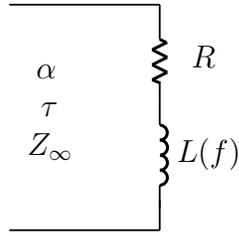
L'induttanza è nota nota i coefficienti $L_{0\div 3}$, ossia i coefficienti dello sviluppo di Taylor del comportamento in frequenza dell'induttanza equivalente. I parametri del modello (sia in questo caso sia nei casi che verranno analizzati in seguito) sono determinabili per esempio mediante una simulazione elettromagnetica dello standard, dalla quale si tirano fuori i parametri scattering, che vengono fittati. In pratica, sono dati da HP, Agilent, Anritsu etc.

Nel modello utilizzato, τ non è una funzione della frequenza f : di fatto, anche se dovrebbe essere così, si tiene conto della dipendenza di τ da f già nell'induttanza.

Osservando su di un software nella carta polare la risposta a perturbazioni sulla fase, si potrebbe vedere che la fase varia parecchio, il modulo no: se non ci sono perdite nella linea, perturbando τ si perturba esclusivamente la fase, come si può immaginare che sia. Quello che capita, perturbando modulo e fase, è il fatto che le varianze deformano in cerchi, ma il fatto che se vi sono covarianze tra fase e modulo può portare ad avere degli ellissi invece che rette o cerchi.

Standard open

Per quanto riguarda lo standard open, c'è relativamente poco da dire: esso è sostanzialmente lo standard duale rispetto allo short:

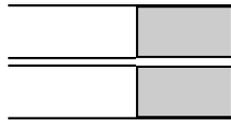


In questo caso sarà la capacità a nascondere la dipendenza dalla frequenza, e dunque a essere sviluppata mediante Taylor:

$$C(f) \sim C_0 + C_1 f + C_2 f^2 + C_3 f^3$$

Il delay invece è circa uguale a quello dello short.

L'open, degli standard, è quello più critico: per esso infatti è necessario, al fine di avere parametri accurati, valutare la radiazione, e ciò non è banale. Nella pratica, l'open si realizza in questa maniera:



In maniera abbastanza simile allo short, non collegando il pin centrale con la calza mediante del dielettrico, ma mettendo del dielettrico: in questa maniera, sostanzialmente, si ha una capacità, ma il campo risulta essere elettromagneticamente confinato. La parte finale, capacitiva, rappresenta il carico dello short. Originariamente, negli anni '80, si introduceva nel maschio uno slab in aria, che carica capacitivamente il termine del coassiale.

Standard load

Lo standard load è piuttosto semplice: esso è semplicemente costituito da una resistenza e da un'induttanza, dove l'induttanza però non è variabile in frequenza (si usa semplicemente un'induttanza del primo ordine). Non si ha una linea di trasmissione, per due motivi: di solito la linea non c'è ma, anche nel caso vi fosse, essa è solitamente ben matchata: la resistenza è sostanzialmente solo l'impedenza caratteristica del coassiale, dunque che ci sia del delay o meno non cambia niente: essendo tutto adattato, non si ha niente di interessante.

Standard reflect

Per reflect si intende uno standard non noto, che si comporta più o meno (a seconda di chi definisce lo standard) come un open o uno short. Il senso fondamentale è: è **non noto** (si può definire, coi type, sia open sia short). Dalle tecniche di taratura è trattato come non noto, ma bisogna definire per esso approssimativamente un ritardo. I due parametri sono dunque il ritardo, e il tipo di impedenza di chiusura. Non importa il valore esatto, ma basta che si abbia, come accuratezza, 180° . Il magnitude del coefficiente di riflessione non è importante, dal momento che esso è ben stimato dalle tecniche di taratura, mentre l'accuratezza con cui va stimata la fase dello standard incognito è 180° . Dalle tecniche di taratura, come vedremo, ciò che esce è la radice di un numero, ma questa può essere nota con una certa indeterminazione.

Standard thru

Lo standard thru è uno standard a due porte: si tratta sostanzialmente di un doppio bipolo completamente noto. Talvolta si differenzia questo standard dallo standard line, ma sostanzialmente essi sono la stessa cosa: uno ha una lunghezza (un ritardo) non nullo (la linea), mentre il thru è sostanzialmente puntiforme, a parametri concentrati. I parametri che permettono di specificarlo sono la lunghezza della linea, il delay, il loss. Lo standard thru (o line) ha implicitamente le impedenze di riferimento a sinistra e a destra; questa cosa è limitante, dal momento che, volendo calibrare un analizzatore di reti con due impedenze diverse, non ci riusciamo; potessimo definire una linea o un thru con due impedenze diverse, sarebbe molto più facile, ma in pratica non c'è nessun analizzatore di reti con uno standard in grado di fare ciò.

Avere uno standard thru significa, in sostanza, avere le due porte dell'analizzatore di rete connesse tra di loro.

Note conclusiva

Abbiamo definito in questa maniera gli standard; essi, sostanzialmente, sono quelli che costituiscono il CALKIT, ossia l'insieme dei coefficienti che devono essere definiti nella memoria dei coefficienti che andranno a pilotare il meccanismo di correzione dell'errore: a seconda della taratura usata e dunque degli standard da usare il CALKIT richiederà i suddetti parametri. È assolutamente fondamentale che i parametri da inserire nel CALKIT siano quelli giusti, forniti dalla ditta produttrice, altrimenti la misura risulterebbe essere completamente errata.

Su un analizzatore di reti, a seconda del tipo di connettore (guida, coassiale), posso caricare diversi CALKIT; è possibile definire altri standard, a seconda delle proprie esigenze o della presenza di fixture.

Per poter parlare di calibrazione, è necessario:

1. partire dalla costruzione di un modello di errore;
2. trovare un algoritmo in grado di determinare i parametri del modello di errore;
3. mediante l'algoritmo, determinare una sequenza di standard, ossia una sequenza di standard da collegare allo strumento al fine di performare la calibrazione.

Questo è l'insieme dei passi necessari al fine di misurare la tara. Si consideri, per esempio, una delle calibrazioni più interessanti (che analizzeremo in seguito): la TRL (Thru, Reflect, Line): si determina una sequenza di passi basata sull'uso di tre standard, ossia un doppio bipolo noto Thru, il doppio bipolo incognito Reflect, e una linea nota Line; in questo caso il thru è a lunghezza nota (solitamente il concetto di thru è da discutere, dal momento che in quasi tutti gli analizzatori di spettro sono assunti a lunghezza 0), però si ha il problema dell'indeterminazione dei piani di riferimento: essi saranno, di fatto, "a metà del thru", che però dovrebbe essere di lunghezza 0, dunque alla fine si potrebbe anche pensare il thru come di lunghezza non nulla; questa cosa è stata problematica per molti modelli come l'8510, 8720 e altri, ma è sostanzialmente un problema del firmware.

Si potrà, dopo ciò, definire il modello di errore.

3.4.4 Definizione del modello di errore

Al fine di definire il modello di errore, primo passo per lo sviluppo di qualunque tecnica di taratura, si faranno due ipotesi fondamentali:

- tutti i componenti del sistema saranno considerati LTI: Lineari Tempo Invarianti;
- si considera il sistema diviso in due parti non comunicanti tra loro: una parte sinistra e una parte destra (questa è un'ipotesi molto forte).

Discutiamo un attimo il discorso della seconda ipotesi: si immagini per esempio di avere due sonde vicine, in un wafer; per esempio, si immagini che esse siano vicine 2 mm. L'ipotesi dice che quando una sonda inietta, dall'altra parte non deve arrivare nulla. Per discutere questa ipotesi, si deve parlare di due tipi di rumore:

- high-level noise: quello che si ottiene quando si misura dalla stessa porta: quando ho un thru, faccio una prima misura alla porta interessata, lo tolgo, faccio una seconda misura, e vedo la variazione dovuta al thru: quanto si trova di differenza facendo la misura con e senza il thru.
- low-level noise: una volta adattate tutte le porte, e introdotto segnale in una, è quello che si vede nelle porte adattate e teoricamente isolate rispetto a quella eccitata; in altre parole, questo è il **crosstalk**.

Il fatto di accettare la seconda ipotesi significa sostanzialmente supporre che né il low-level né l'high-level noise esistano. Ciò che si fa è isolare completamente le due metà, e fingere che non parlino per niente.

Si ha a che fare, con questi problemi, con 8 incognite, per ciascuna delle metà: a_0, b_0, a_1, b_1 (o a_2, b_2 nella seconda metà), a_3, b_3, a_4, b_4 . A ogni porta inoltre ho ciò:

Questi non sono le uniche relazioni che permettono di definire il problema: la catena di downconversion infatti ci permette di ottenere altre equazioni, dal momento che la catena dà un segnale V , misurato alla porta 1, che è uguale a un certo K per b_3 :

Si noti che il fatto che il coefficiente K moltiplichi V_3 o b_3 non cambia nulla: se non cambiamo il coefficiente di riflessione Γ_3 , la tensione totale misurata è

$$V_3 = (1 + \Gamma_3)b_3$$

dunque, non cambiando la riflessione, il legame tra tensione totale e onda di potenza è piuttosto stretto.

Cosa abbiamo dunque in mano?

- la matrice scattering del quattro porte: essa è la matrice scattering non solo dell'accoppiatore direzionale, ma del quattro porte di cui due sono caricate sui mixer, una sulla sorgente, l'ultima è il piano di riferimento;
- le equazioni di carico del mixer;
- le equazioni di conversione.

Partiamo dunque dalla matrice scattering del quattro porte: essa sarà definibile a partire da una forma del tipo:

$$\begin{bmatrix} b_0 \\ b_1 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} \\ S_{21} & S_{22} & S_{23} & S_{24} \\ S_{31} & S_{32} & S_{33} & S_{34} \\ S_{41} & S_{42} & S_{43} & S_{44} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_3 \\ a_4 \end{bmatrix}$$

Si noti che nella definizione non sono presenti le onde relative alla porta 2; questo è dovuto al fatto che essa è “nell’altra metà del dispositivo”, dunque lontano.

Avendo a che fare con una relazione matriciale, essa è sostanzialmente lineare; il fatto che il sistema sia LTI implicherà dunque sicuramente il fatto che tutte le relazioni presenti continuino a essere lineari, e dunque abbiano sostanzialmente un aspetto matriciale. Date V_{m1} e V_{m2} tensioni totali misurate dai due mixer, ciò che stiamo cercando è una relazione del tipo:

$$\begin{cases} a_1 = c_1 V_{m1} + c_2 V_{m2} \\ a_2 = c_3 V_{m1} + c_4 V_{m2} \end{cases}$$

ossia, due combinazioni lineari pesate con i coefficienti c_i . Per il nostro problema, si considerano le due tensioni totali come le variabili indipendenti del problema. Si ha ancora, come equazioni costitutive:

$$\begin{cases} a_3 = \Gamma_3 b_3 \\ a_4 = \Gamma_4 b_4 \end{cases}$$

e le relazioni di carico dei mixer:

$$\begin{cases} V_{m1} = K_3 b_3 \\ V_{m2} = K_4 b_4 \end{cases}$$

Scriviamo dunque la matrice scattering in forma di sistema:

$$\begin{cases} b_0 = S_{11}a_0 + S_{12}a_1 + S_{13}a_3 + S_{14}a_4 \\ b_1 = S_{21}a_0 + S_{22}a_1 + S_{23}a_3 + S_{24}a_4 \\ b_3 = S_{31}a_0 + S_{32}a_1 + S_{33}a_3 + S_{34}a_4 \\ b_4 = S_{41}a_0 + S_{42}a_1 + S_{43}a_3 + S_{44}a_4 \end{cases}$$

Sostituiamo, in questa, le equazioni costitutive:

$$\begin{cases} b_0 = S_{11}a_0 + S_{12}a_1 + S_{13}\Gamma_3 b_3 + S_{14}\Gamma_4 b_4 \\ b_1 = S_{21}a_0 + S_{22}a_1 + S_{23}\Gamma_3 b_3 + S_{24}\Gamma_4 b_4 \\ b_3 = S_{31}a_0 + S_{32}a_1 + S_{33}\Gamma_3 b_3 + S_{34}\Gamma_4 b_4 \\ b_4 = S_{41}a_0 + S_{42}a_1 + S_{43}\Gamma_3 b_3 + S_{44}\Gamma_4 b_4 \end{cases}$$

Ordino, ponendo al membro sinistro i termini in a_0, b_0, a_1, b_1 , a destra gli altri; ottengo:

$$\begin{cases} -S_{11}a_0 - S_{12}a_1 + b_0 = S_{13}\Gamma_3 b_3 + S_{14}\Gamma_4 b_4 \\ -S_{21}a_0 - S_{22}a_1 + b_1 = S_{23}\Gamma_3 b_3 + S_{24}\Gamma_4 b_4 \\ -S_{31}a_0 - S_{32}a_1 = -b_3 + S_{33}\Gamma_3 b_3 + S_{34}\Gamma_4 b_4 \\ -S_{41}a_0 - S_{42}a_1 = -b_4 + S_{43}\Gamma_3 b_3 + S_{44}\Gamma_4 b_4 \end{cases}$$

In forma matriciale, questo diventa semplicemente:

$$\begin{bmatrix} -S_{11} & 1 & -S_{12} & 0 \\ -S_{21} & 0 & -S_{22} & 1 \\ -S_{31} & 0 & -S_{32} & 0 \\ -S_{41} & 0 & -S_{42} & 0 \end{bmatrix} \begin{bmatrix} a_0 \\ b_0 \\ a_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} S_{13}\Gamma_3 & S_{14}\Gamma_4 \\ S_{23}\Gamma_3 & S_{24}\Gamma_4 \\ S_{33}\Gamma_3 - 1 & S_{34}\Gamma_4 \\ S_{43}\Gamma_3 & S_{44}\Gamma_4 - 1 \end{bmatrix} \begin{bmatrix} b_3 \\ b_4 \end{bmatrix}$$

Questo può essere scritto in maniera più compatta, nella seguente maniera:

$$\underline{\underline{W}} \begin{bmatrix} a_0 \\ b_0 \\ a_1 \\ b_1 \end{bmatrix} = \underline{\underline{Q}} \begin{bmatrix} b_3 \\ b_4 \end{bmatrix}$$

Non è finito: a questo punto, è necessario legare questa equazione alle grandezze misurabili, ossia alle tensioni totali che si ottengono in uscita dai mixer; per fare ciò è necessario utilizzare le equazioni di carico dei mixer. In particolare alla porta 3 si misura l'onda progressiva alla porta 1 del DUT (che verrà chiamata a_{m1}), mentre alla porta 4 si misura l'onda regressiva alla porta 2 del DUT (dunque b_{m1}). Si noti inoltre che ciò che effettua le misure, come già detto, è un voltmetro vettoriale, e esso è in grado di misurare non un'onda di potenza, ma una tensione totale; è necessario dunque tenere conto di ciò, al momento di definire i parametri K_i . Si ha che:

$$a_{m1} = b_3(1 + \Gamma_3)\sqrt{Z_{r3}} \triangleq K_3 b_3$$

$$b_{m1} = b_4(1 + \Gamma_4)\sqrt{Z_{r4}} \triangleq K_4 b_4$$

o, col formalismo matriciale:

$$\begin{bmatrix} a_{m1} \\ b_{m1} \end{bmatrix} = \begin{bmatrix} K_3 & 0 \\ 0 & K_4 \end{bmatrix} \begin{bmatrix} b_3 \\ b_4 \end{bmatrix} \triangleq \underline{\underline{K}} \begin{bmatrix} b_3 \\ b_4 \end{bmatrix}$$

Inserendo questa equazione nelle precedenti, invertendo, si ottiene semplicemente:

$$\begin{bmatrix} a_0 \\ b_0 \\ a_1 \\ b_1 \end{bmatrix} = \underline{\underline{W}}^{-1} \underline{\underline{Q}} \underline{\underline{K}}^{-1} \begin{bmatrix} a_{m1} \\ b_{m1} \end{bmatrix} \triangleq \underline{\underline{D}} \begin{bmatrix} a_{m1} \\ b_{m1} \end{bmatrix}$$

Dove a_{m1} e b_{m1} sono ciò che si misura: questi sono i parametri che si trovano in uscita dal ADC, dunque essi sono solamente numeri, privi di

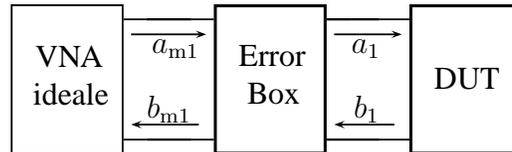
significato fisico (per quanto collegati, come detto prima, alle onde di potenza). L'equazione matriciale, considerando i coefficienti D_{ij} della matrice, può essere scritta come:

$$\begin{cases} a_0 = D_{11}a_{m1} + D_{12}b_{m1} \\ b_0 = D_{21}a_{m1} + D_{22}b_{m1} \\ a_1 = D_{31}a_{m1} + D_{32}b_{m1} \\ b_1 = D_{41}a_{m1} + D_{42}b_{m1} \end{cases}$$

Le ultime due righe sono quelle che ci interessano: esse sostanzialmente permettono di identificare la relazione tra le onde di potenza misurate e ciò che entra alla porta 1 del sistema. Questa ultima relazione, dunque, può essere scritta in una forma particolare, che ricorda sostanzialmente una matrice scattering: essa in realtà non è una vera matrice scattering, dal momento che, per quanto dimensionalmente potrebbe ricordarla, non rispetta le proprietà della rappresentazione scattering (reciprocità, unitarietà in caso di assenza di perdite e simili). Ciò che si trova, dunque, alla fine di tutto, mescolando queste ultime relazioni, è la cosiddetta **error box**:

$$\begin{cases} b_{m1} = e_{11}a_{m1} + e_{12}b_1 \\ a_1 = e_{21}a_{m1} + e_{22}b_1 \end{cases}$$

questa è stata ottenuta semplicemente rimescolando le ultime due righe della matrice precedente. Ciò permette di ottenere un legame tra i parametri fisici e quelli numerici misurati. L'ultima matrice, invece, è necessaria al fine di esprimere il sistema nella seguente maniera:



Un insieme di due blocchi (+ il DUT): un analizzatore di reti ideale, che misura correttamente e senza problemi a_{m1} e b_{m1} , e una error box, ossia una “scatola” che al proprio interno contiene tutte le informazioni sugli errori. La matrice degli e_{ij} è quella che rappresenta il sistema come da figura. Si noti che la error box è un modello, ma di fatto non vi è nessun analizzatore di reti ideale o circuito che rappresenti la error box: si tratta solo di un modello matematico che permette di legare i parametri di un dispositivo ai parametri misurati.

Un'osservazione fondamentale: date le equazioni della error box, i parametri della error box dipendono dai valori di sorgente? Ossia, se si usa un analizzatore di reti con la solita configurazione, i parametri cambiano variando

l'ingresso? La risposta è no: infatti, come si può vedere, se si partisse dalle equazioni iniziali, non si potrebbe capire; mediante la nostra derivazione, si vede chiaramente che i parametri che riguardano la sorgente, a_0 e b_0 , non sono all'interno delle nostre relazioni: nelle nostre equazioni non ci sono le condizioni di carico e sorgente! Questo fatto è dovuto all'ipotesi di linearità del circuito: le informazioni sul generatore a radiofrequenza è contenuta in a_0 e b_0 , ma nelle equazioni trovate non si ha apparente traccia di informazioni sul generatore. Questo succede perchè se cambiamo il generatore, di sicuro a_1 e b_1 variano, ma variano in maniera proporzionale anche le misure ottenute, a_{m1} e b_{m1} , e questo comportamento è tipico dei comportamenti lineari. Questa cosa ha conseguenze enormi: se tolgo una sorgente, e ne metto un'altra, posso usare comunque la calibrazione fatta con la sorgente precedente, e questa cosa è importantissima: facilita notevolmente le cose rispetto al caso pessimistico.

Tutto il discorso fatto alla porta 1, quella di ingresso, vale anche alla porta 2: come accennato, qualsiasi sia la terminazione, il carico, la calibrazione continua a essere sempre giusta. Questo significa che posso sostituire la terminazione alla porta 2 (di carico dunque) con una più potente (se devo fare misure di potenza), ed effettuare la misura, o comunque diversa (cambiare per esempio, dopo la taratura, gli accoppiatori direzionali, senza dover ritarare). Questo per esempio è fatto per misurare i parametri dei transistori di potenza.

L'approccio utilizzato è sistemico: non si è detto, di fatto, come è fatta la matrice di scattering del sistema; questo significa che i parametri sono definiti a meno della conoscenza della matrice scattering del riflettometro, non conosciamo i K_i : per poter identificare questi coefficienti è necessario dunque passare per la tecnica di taratura, che ci permetterà di determinarli.

Il caso più bello che ci potrebbe capitare è:

- impedenza di riferimento circa uguale a quella del riflettometro;
- accoppiatori direzionali sostanzialmente ideali;
- strumento di misura perfetto.

Ciò che si cercava di fare è prendere un accoppiatore direzionale il più possibile ideale, impedenze di riferimento tali da avere adattamento, quella con cui l'accoppiatore era costruito, e per il voltmetro si può dimostrare che, dal momento che si misurano i rapporti, anche se $K \neq 1$ non si hanno problemi. Si può dimostrare che, nel caso in cui le condizioni siano rispettate, si ha che:

- $e_{11} \sim D$, dove D è la direttività dell'accoppiatore direzionale;
- $e_{22} \sim Z_{\text{source match}}$, dove per source match si intende l'impedenza che si vede quando si guarda dentro l'accoppiatore direzionale;
- $e_{12}e_{21}$ sono circa uguali al tracking dell'accoppiatore direzionale.

Si hanno tre parametri, non quattro, nella nostra formula: al fine di effettuare la taratura, sostanzialmente, si deve trovare, come parametro, il rapporto a_2/b_2 , e questo si può scrivere come:

$$\Gamma_m = e_{11} + \frac{e_{12}e_{21}\Gamma_x}{1 - e_{22}\Gamma_x}$$

dove Γ_m è il gamma misurato, e Γ_x il coefficiente di riflessione opposto dal carico. Questo può essere scritto, nel caso si abbiano le condizioni di prima:

$$\Gamma_m \sim E_D + \frac{E_T\Gamma_x}{1 - E_S\Gamma_x}$$

dove per E_D si intende l'errore di direttività, ossia e_{11} , per E_T l'errore di tracking, ossia $e_{12}e_{21}$, E_S l'errore di source matching, ossia e_{22} . I parametri utili dunque sono 3 e non 4.

Tutti questi ragionamenti, si vuole ribadire, valgono a un'ipotesi: si sta parlando di porta 2, dunque di terminazioni. Se io cambiassi però la sorgente, e con essa l'impedenza di sorgente, sembra significare che cambio il source match. Questo è un errore: la formula di taratura vale solo e solamente se si ha adattamento di tutte le formule; cambiare dunque i carichi o le sorgenti e parlare ancora di source match nella formula di riferimento non ha senso, dal momento che solo se tutto è adattato la formula vale. Ciò che si deve fare è, nota l'impedenza di accoppiatore, cavi e tutto il resto, fare la taratura con questa impedenza, e dunque avere il permesso di utilizzare la suddetta formula di taratura. In verità dunque si parla di source match, ma non si ha nessun problema per il source: dopo la taratura possiamo pure cambiarlo.

Conviene dunque effettuare il procedimento di taratura con tutto adattato dal momento che, se si fa il calcolo della sensibilità, ossia della variazione dei coefficienti di taratura (CALKIT), si vede che se si tara con un'altra impedenza si hanno coefficienti molto diversi da quelli del caso ideale: il caso ideale ha $E_D = E_S = 0$, $E_T = 1$, dunque una error box di forma

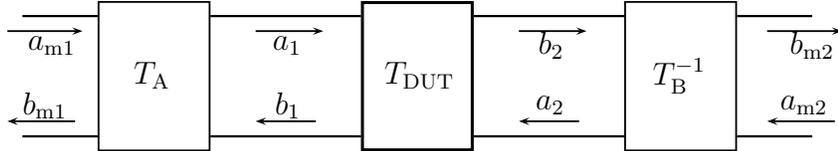
$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

antidiagonale. Se l'impedenza di riferimento usata dagli standard è dunque uguale a quella del sistema fisico, la sorgente usata per la taratura idem, e

così via, si ottiene qualcosa di prossimo a ciò, e dunque gli errori saranno corretti efficacemente.

3.4.5 Modello di errore a due porte

Se il modello in questione è di un 2-porte, allora si deve estendere il modello nella seguente maniera:



Il fatto che si abbia una \underline{T}^{-1} è storico: il primo articolo che definì in maniera formale il problema del modello di errore di 2-porte utilizzò questa notazione, che è rimasta poi radicata.

Abbiamo un problema: vogliamo fare la misura della matrice scattering del 2-porte, ma dunque vogliamo scrivere qualcosa del tipo:

$$\begin{bmatrix} b_{m1} \\ b_{m2} \end{bmatrix} = \begin{bmatrix} S_{m,11} & S_{m,12} \\ S_{m,21} & S_{m,22} \end{bmatrix} \begin{bmatrix} a_{m1} \\ a_{m2} \end{bmatrix}$$

Il nostro strumento ideale, in questo modello dove si separa l'error box (o, in questo caso, i due error box) dall'analizzatore ideale; avremmo:

$$\begin{bmatrix} b_{m1} \\ b_{m2} \end{bmatrix} = \underline{S}_m \begin{bmatrix} a_{m1} \\ a_{m2} \end{bmatrix}$$

L'idea sarebbe: misuriamo, con l'analizzatore di reti ideale, \underline{S}_m , quindi, supponendo di conoscere le matrici \underline{T}_A e \underline{T}_B^{-1} , si potrebbe trovare la misura. Purtroppo ciò non si può fare, dal momento che per misurare la matrice scattering ideale dovremmo essere in grado di "adattare le porte dell'analizzatore di reti ideale", e ciò non ha senso: noi non abbiamo un analizzatore di reti ideale, dal momento che esso è solo un modello matematico.

Ciò che si fa in pratica per trattare il caso della misura di 2-porte è qualcosa di diverso: il nostro obiettivo è ottenere un certo insieme di misure al fine di poter rendere il sistema ben determinato, ma dunque possiamo misurare in due condizioni; una volta misurate le matrici scattering in due condizioni, si può semplicemente dire che:

$$\begin{bmatrix} S_{m,11} & S_{m,12} \\ S_{m,21} & S_{m,22} \end{bmatrix} = \begin{bmatrix} b'_{m1} & b''_{m1} \\ b'_{m2} & b''_{m2} \end{bmatrix} \begin{bmatrix} a'_{m1} & a''_{m1} \\ a'_{m2} & a''_{m2} \end{bmatrix}^{-1}$$

ossia misuro i parametri b_{mi}/a_{mi} in due condizioni di misura. Le due matrici a secondo membro non sono matrici scattering, ma semplicemente delle tabelle con i numeri misurati, in modo coerente tra loro (per ottenere i singoli parametri). Ciò che si fa in pratica è

$$\underline{S}_m = \underline{b}_m \underline{a}_m^{-1}$$

Ossia faccio altre 4 misure, al fine di avere più equazioni.

Come si può fare un'altra misura? Sostanzialmente, così:

Lo switch può essere posizionato, come visto, in due modi; semplicemente si potranno fare le misure in due posizioni, e così possiamo fare 4 misure extra semplicemente commutando lo switch.

Quello che stiamo considerando è un modello a 8 termini, anche se in realtà di termini interessanti per davvero ve ne sono solo 7: di onde, in assoluto, non se ne hanno 8, dal momento che ci interessa la matrice \underline{T} del DUT, e questa si trova a meno di un termine. Vedremo nel dettaglio.

3.4.6 Principali tecniche di taratura

In questa sezione a questo punto si introdurranno, almeno in maniera qualitativa, le principali tecniche di taratura utilizzate.

TSD e TRL: Thru, Short, Delay e Thru, Reflect, Line

Questa tecnica esiste in più forme: essa è stata scoperta nel '75 circa contemporaneamente negli Stati Uniti e a Genova, ma quella che noi studieremo è una variante a una versione ancora successiva, del '85. La tecnica di taratura in questione si basa sull'inserzione di un thru, ossia di un doppio bipolo completamente noto, una linea, ossia un doppio bipolo di cui è nota sola l'impedenza caratteristica, che oltre a essere nota deve essere fisicamente uguale a quella del thru; l'algoritmo deve essere dunque in grado di ricavare il βl della linea. A seconda che si abbia o meno lo standard a una porta noto o incognito, si hanno rispettivamente TSD o TRL. La cosa che differenzia le due tarature dunque è il fatto che il terzo standard (dopo thru e line) sia noto (TSD) o incognito (TRL).

Questa tecnica è quella che permette di ottenere i risultati migliori, ossia le misure più accurate, dal momento che essa è basata su di una linea, dunque su standard distribuiti, dunque standard a microonde; questa taratura è quella che si utilizza dunque nei laboratori di metrologia. Il grosso problema della TRL è che bisogna fare tutto questo lavoro di taratura ai piani di riferimento, spostare le probe, e ciò su di un wafer di silicio non si può fare: a livello di produzione non ha senso. La TRL, come vedremo, ha inoltre problemi

dovuti alla risonanza della linee: quando la linea risuona, dal momento che $e^{-jkl} = 1$ per qualche frequenza e dunque per qualche k , le varie reattanze del thru e della linea vanno a compensarsi, facendo decadere il funzionamento del sistema: la TRL è band limited.

Come funziona in parole povere? Servono, come noto, 7 equazioni; abbiamo ciò:

- il thru ha quattro parametri noti, dunque quattro equazioni;
- la linea dà altre quattro equazioni;
- la reflection, non nota, misurata uguale alle due porte, darà due equazioni;

in totale si hanno 10 equazioni, ne servirebbero 7, per rendere il sistema determinato, ma poi si ha un termine incognito della linea, uno incognito della reflection. Si può dimostrare che il set di equazioni dato dal thru e dalla linea dà però luogo a sole sei equazioni indipendenti.

LRM: Line Reflect Match

In questo tipo di taratura, arrivata circa 15 anni dopo le prime pubblicazioni (si parla del '90), con un articolo che, tra le altre cose, formalizza l'approccio sistemico, si fa ciò: si introduce una linea, ossia un thru con ritardo, un reflect (standard incognito alla porta 1 e 2), e un match alle due porte. Il fatto di usare questo approccio evita di far spostare le probe, ottenendo un metodo band unlimited: il match a 50Ω è a larga banda. Esiste la LRRM, che usa due reflection, con le quali si misura anche l'induttanza parassita, in modo da migliorare ulteriormente l'accuratezza.

Come funziona in parole povere?

- la line ha quattro parametri noti (i quattro parametri della matrice scattering), dunque quattro equazioni;
- il match, misurando le due porte, darà due parametri;
- la reflection, non nota, misurata uguale alle due porte, darà due equazioni;

servono 7 parametri, più uno incognito della reflection, e in totale ho 8 equazioni conosciute.

SOLR : Short, Open, Load, Reciprocal

Fino al 1991 tutte le tecniche richiedevano almeno un doppio bipolo completamente conosciuto: il thru. Questo significava che, in qualche modo, era sempre necessario collegare le due porte; esistono molti casi (per esempio quando si hanno due connettori dello stesso sesso), dove creare questo collegamento è duro. In queste situazioni, dunque, si può utilizzare una taratura di tipo diverso. Nel 1995 si definì la SOLR: essa permette di sostituire il thru con un oggetto reciproco.

In questo caso si hanno tre standard alla porta 1 tutti noti (short, open, load), idem per la porta 2, e la condizione di reciprocità del reciprocal dà l'ultima equazione utile per arrivare a 7.

SOLT : Short, Open, Load, Thru

Questa taratura, “tradizionale”, si basa su di uno Short, un Open, un Load, un Thru: questa è fondamentale e obbligatoria quando si ha a che fare con dei VNA a 3 sampler. Questa, se utilizzata in un 4-sampler, è una taratura sovradeterminata: del thru sono interessato solo, nel caso dei 4-sampler, a una caratteristica: la sua reciprocità. Se in questo caso ho nozioni anche sul thru, allora ho più dati di quelli che effettivamente mi servono: usando questo set di standard è possibile risolvere il sistema di taratura con i minimi quadrati, e fare una stima dei coefficienti di errore diversa dalla solita. Se le prime tre in un 4-sampler risolvono il problema “one shot”, senza dover sovradeterminare, questa deve richiedere una stima di tipo diverso. Ho sei standard come prima, più il thru che ne dà altri 4 e permette di arrivare a dieci.

3.5 Tecniche di taratura

3.5.1 Approccio basato sulle matrici di trasmissione

Come già detto in precedenza, le tarature sono sostanzialmente basate sull'uso di una sequenza di standard. Si parta dalla seguente equazione matriciale, detta “equazione di taratura per sistemi a due porte”:

$$\underline{T}_{\text{standard},M} = \underline{T}_A \underline{T}_{\text{standard}} \underline{T}_B^{-1}$$

Cosa significa ciò? Data la matrice dei parametri di un generico standard, e la misura dei parametri dello standard (i numeri raw), è possibile legare mediante le due matrici \underline{T}_A e \underline{T}_B .

Applichiamo un thru: l'equazione matriciale in questo caso sarà:

$$\underline{\underline{T}}_{\underline{\underline{MT}}} = \underline{\underline{T}}_{\underline{\underline{A}}}\underline{\underline{T}}_{\underline{\underline{T}}}\underline{\underline{T}}_{\underline{\underline{B}}}^{-1}$$

Applichiamo un delay, ossia una linea; ora, in questo caso, l'equazione matriciale sarà:

$$\underline{\underline{T}}_{\underline{\underline{ML}}} = \underline{\underline{T}}_{\underline{\underline{A}}}\underline{\underline{T}}_{\underline{\underline{L}}}\underline{\underline{T}}_{\underline{\underline{B}}}^{-1}$$

Calcoliamo a questo punto una matrice $\underline{\underline{R}}_{\underline{\underline{M}}}$ definita come:

$$\underline{\underline{R}}_{\underline{\underline{M}}} \triangleq \underline{\underline{T}}_{\underline{\underline{ML}}}\underline{\underline{T}}_{\underline{\underline{MT}}}^{-1}$$

Mediante passaggi algebrici matriciali, si trova banalmente che:

$$\begin{aligned} \underline{\underline{R}}_{\underline{\underline{M}}} &\triangleq \underline{\underline{T}}_{\underline{\underline{ML}}}\underline{\underline{T}}_{\underline{\underline{MT}}}^{-1} = \underline{\underline{T}}_{\underline{\underline{A}}}\underline{\underline{T}}_{\underline{\underline{L}}}\underline{\underline{T}}_{\underline{\underline{B}}}^{-1} \left(\underline{\underline{T}}_{\underline{\underline{A}}}\underline{\underline{T}}_{\underline{\underline{T}}}\underline{\underline{T}}_{\underline{\underline{B}}}^{-1} \right)^{-1} = \\ &= \underline{\underline{T}}_{\underline{\underline{A}}}\underline{\underline{T}}_{\underline{\underline{T}}}\underline{\underline{T}}_{\underline{\underline{B}}}^{-1} \underline{\underline{T}}_{\underline{\underline{B}}}\underline{\underline{T}}_{\underline{\underline{T}}}\underline{\underline{T}}_{\underline{\underline{A}}}^{-1} = \underline{\underline{T}}_{\underline{\underline{A}}}\underline{\underline{T}}_{\underline{\underline{L}}}\underline{\underline{T}}_{\underline{\underline{T}}}\underline{\underline{T}}_{\underline{\underline{A}}}^{-1} \end{aligned}$$

Vediamo a questo punto un paio di osservazioni: come noto, la matrice scattering di una linea di trasmissione è la seguente:

$$\underline{\underline{S}}_{\underline{\underline{linea}}} = \begin{bmatrix} 0 & e^{-jkl_L} \\ e^{-jkl_L} & 0 \end{bmatrix}$$

$$\underline{\underline{S}}_{\underline{\underline{thru}}} = \begin{bmatrix} 0 & e^{-jkl_T} \\ e^{-jkl_T} & 0 \end{bmatrix}$$

Infatti, di fatto il delay è una linea, e il thru anche (può essere una linea di lunghezza 0); l_L è la lunghezza della linea, l_T la lunghezza del thru. Dalla Teoria delle Reti, è noto il fatto che una generica matrice di trasmissione (matrice utilizzata per la rappresentazione del modello a due porte) è trasformabile nel seguente modo a partire dai parametri scattering:

$$\underline{\underline{T}} = \begin{bmatrix} \frac{\Delta_{\underline{\underline{S}}}}{S_{21}} & \frac{S_{11}}{S_{21}} \\ -\frac{S_{22}}{S_{21}} & \frac{1}{S_{21}} \end{bmatrix}$$

dove $\Delta_{\underline{\underline{S}}} = S_{11}S_{22} - S_{12}S_{21}$.

Da ciò, si può calcolare che:

$$\underline{\underline{T}}_{\underline{\underline{L}}}\underline{\underline{T}}_{\underline{\underline{T}}}^{-1} = \begin{bmatrix} e^{jk(l_L - l_T)} & 0 \\ 0 & e^{-jk(l_L - l_T)} \end{bmatrix}$$

Osservazione interessante: si può dire che, rispetto a $\underline{\underline{R}}_{\underline{\underline{ML}}}$, la matrice $\underline{\underline{T}}_{\underline{\underline{A}}}$ sia la sua matrice di cambio di base: la matrice $\underline{\underline{T}}_{\underline{\underline{L}}}\underline{\underline{T}}_{\underline{\underline{T}}}^{-1}$ infatti è una matrice

diagonale, dunque si può dire che il fatto che sia moltiplicata a sinistra per una certa matrice e a destra per la sua inversa implica che questa certa matrice, $\underline{\underline{T}}_A$, sia la sua matrice diagonalizzante.

Prima di continuare i discorsi, è necessario ribadire un fatto: l'ipotesi che sottostava al poter scrivere tutto ciò, a una porta, era il fatto che l'impedenza della linea coincidesse con l'impedenza di riferimento utilizzata. Nel caso a due porte, questo fatto non è fondamentale: l'impedenza di riferimento scelta può essere arbitraria, ma si ha una condizione molto importante: quando si ha a che fare con due matrici, l'ipotesi fondamentale è il fatto che linea e thru abbiano le stesse impedenze **caratteristiche**, e non **di riferimento**. L'impedenza caratteristica è una definizione legata alla fisica del circuito, alla geometria della struttura guidante, nella fattispecie alla sua sezione trasversale; l'impedenza di riferimento è un'impedenza che si utilizza semplicemente per calcolare i parametri scattering, dunque un qualcosa privo di significato fisico. Nel caso (che si trova) in cui il thru è una semplice connessione, in cui dunque non si ha lunghezza del thru, l'impedenza caratteristica della linea è quella che andrà a diventare anche l'impedenza di riferimento delle porte dello strumento. Si tenga inoltre conto che, nelle nostre trattazioni, non si considerano effetti secondari quali la presenza di modi superiori o di perdite: tutto è monomodale e privo di perdite. Come vedremo, la TRL è una tecnica che calibra l'analizzatore di reti, dandole come impedenza di riferimento un'impedenza pari all'impedenza caratteristica della linea.

Ancora qualche osservazione: $\underline{\underline{T}}_{ML}$ è una matrice di numeri, dunque tendenzialmente priva di significato fisico; da essa però si dice che $\underline{\underline{T}}_L \underline{\underline{T}}_T^{-1}$ sia diagonale, e ciò produce tutte le conseguenze viste. È proprio necessario che quest'ultima matrice sia diagonale? Beh, di fatto no: la condizione che ci serve è il fatto che thru e linea di delay abbiano la stessa impedenza caratteristica, non di riferimento; ciò che si deve fare è diagonalizzare la matrice del thru o della linea, in modo da variare la sua impedenza di riferimento e vedere che tutto funziona. Si immagini che thru e linea abbiano impedenze caratteristiche pari a 100 Ω , ma che si abbiano, per esse, una matrice $\underline{\underline{T}}_T^{50}$ e $\underline{\underline{T}}_L^{50}$, ossia matrici di trasmissione calcolate rispetto a carichi di 50 Ω . Cosa si può fare? Beh, come appena detto, diagonalizzare in modo da trovare una matrice diagonale normalizzata all'impedenza caratteristica, 100 Ω ; data $\underline{\underline{A}}$ la matrice di normalizzazione:

$$\underline{\underline{T}}_L^{50} \underline{\underline{T}}_T^{-1,50} = \underline{\underline{A}} \underline{\underline{T}}_L^{100} \underline{\underline{A}}^{-1} \underline{\underline{A}} \underline{\underline{T}}_L^{-1,100} \underline{\underline{A}}^{-1}$$

ciò produce la stessa cosa di prima.

Si torni dunque all'equazione trovata in precedenza:

$$\underline{\underline{R}}_M = \underline{\underline{T}}_A \underline{\underline{T}}_L \underline{\underline{T}}_T^{-1} \underline{\underline{T}}_A^{-1} \triangleq \underline{\underline{T}}_A \underline{\underline{\lambda}} \underline{\underline{T}}_A^{-1}$$

Immaginiamo di avere dunque $\underline{\underline{R}}_M$: se la costruisco e ne calcolo gli autovalori, trovo $\underline{\underline{\lambda}}$. Essa, come visto prima, permette di calcolare senza ancora avere calibrato (di fatto non abbiamo ancora calibrato lo strumento), in modo estremamente accurato, la costante di propagazione k : abbiamo infatti visto prima che la matrice è una matrice diagonale con solo termini di sfasamento; questi termini di sfasamento avranno una lunghezza nota come parametro, dal momento che thru e delay sono doppi bipoli noti, quindi abbiamo un side-effect molto positivo. Possiamo capire quanto accuratamente stiamo inoltre facendo le procedure: dalla matematica/fisica sappiamo che gli autovalori devono essere l'uno il coniugato (o il reciproco) dell'altro, ma, a causa di imperfezioni nella misura, essi non saranno esattamente reciproci. La “distanza” tra i due ci permetterà di capire di quanto stiamo sbagliando le misure.

Per le strutture guidanti, servono sostanzialmente due costanti secondarie: la costante di propagazione, che abbiamo trovato in maniera piuttosto semplice, e l'impedenza caratteristica. Conoscere quest'ultima non è banale, dal momento che essa non ci viene data gratuitamente dalle misure senza calibrazione come l'altra costante secondaria: essa deve dunque essere nota, per qualche via, a priori. Un primo metodo per determinare l'impedenza caratteristica potrebbe essere basato sull'approssimazione quasi-statica: come noto, si ha che

$$Z_\infty = \sqrt{\frac{\mathcal{L}}{\mathcal{C}}}$$

dove

$$\mathcal{C} = \varepsilon_{\text{eff}} \mathcal{C}_{\text{vuoto}}$$

Nota la costante di propagazione k è possibile ricavare la velocità di fase v_f della struttura, dunque un'idea (quella dell'approssimazione quasi-statica) è quella di misurare \mathcal{C}_d come:

$$\mathcal{C}_d \sim \mathcal{C}|_{f=0}$$

dunque, a partire dalla formula

$$v_f = \frac{1}{\sqrt{\mathcal{L}\mathcal{C}}}$$

ricavare \mathcal{L} e dunque la Z_∞ . Un'alternativa potrebbe essere quella di ricavare \mathcal{C}_0 mediante un simulatore elettromagnetico, la ε_{eff} mediante misure, dunque moltiplicare e trovare allo stesso modo Z_∞ .

Rimembrando le lezioni di Geometria, se $\underline{\lambda}$ è la matrice diagonalizzata, allora $\underline{\underline{T}}_A$, a rigor di logica, sarà la matrice degli autovettori:

$$\underline{\underline{T}}_A = [\underline{v}_1 \ \underline{v}_2]$$

Come si dovrebbe fare per fare i conti? Beh, gli autovettori sono, per una matrice $\underline{\underline{A}}$, quei vettori tali per cui si verifica la relazione

$$\underline{\underline{A}}v = \lambda v$$

questi si ottengono imponendo, nelle soluzioni dell'equazione matriciale, un coefficiente a 1, e l'altro avrà un valore ad esso normalizzato. Il fatto di avere un autospazio di dimensione 2 ci permette di avere due gradi di libertà all'interno del nostro sistema: a e b , ossia due coefficienti a meno dei quali è possibile definire gli autovettori.

A patto di utilizzare, in qualsiasi forma noi vogliamo, due soli gradi di libertà, è dunque possibile introdurre una forma in grado di rappresentare la matrice degli autovettori. Come noto dal corso di Algebra Lineare, dunque, per non perdere di generalità, per il primo autovettore si utilizza, come parametro, un certo k , per il secondo un certo p ; dalla soluzione del sistema che permette di determinare il vettore dell'autospazio, si troverà che, se un elemento è un generico k , l'altro sarà ancora k , moltiplicato per un certo a , derivante dalla soluzione per l'appunto del problema:

$$\underline{v}_1 = \begin{bmatrix} ka \\ k \end{bmatrix}$$

allo stesso modo, per il secondo autovettore dell'autospazio, si avrà dipendenza da un certo parametro p :

$$\underline{v}_2 = \begin{bmatrix} pb \\ p \end{bmatrix}$$

dunque, la matrice $\underline{\underline{T}}_A$ sarà:

$$\underline{\underline{T}}_A = \begin{bmatrix} ka & pb \\ k & p \end{bmatrix}$$

Nel caso della matrice degli autovettori relativi alla matrice $\underline{\underline{R}}_M$, dunque, ossia $\underline{\underline{T}}_A$, si sceglie di utilizzare una forma di questo tipo:

$$\underline{T}_A = p \begin{bmatrix} \frac{k}{p}a & b \\ \frac{k}{p} & 1 \end{bmatrix} \triangleq p\underline{X}_A$$

a e b sono i parametri che derivano dal calcolo degli autovettori, dunque sono parametri noti; il problema della taratura sarà risolto una volta calcolato $\frac{k}{p}$. Di tutte le possibili matrici ce ne serve una specifica, dipendente da un certo valore di k e da un certo valore di p : siamo interessati a una specifica matrice, quella che permette di fare la taratura, dunque essa dovrà essere interamente nota, non funzione di un parametro; se ciò che permette di relazionare il parametro all'altro elemento del vettore (in altre parole, a e b) si risolve semplicemente con calcoli di Algebra Lineare, si dovranno fare considerazioni extra per quanto riguarda gli altri elementi, ossia k e p (o meglio, il loro rapporto). Come si può fare questo calcolo? Beh, ricordiamo le nozioni precedentemente ricavate per quanto riguarda l'error box; si ha un'equazione del tipo:

$$\Gamma_{M1} = e_{00} + \frac{e_{10}e_{01}\Gamma_x}{1 - e_{11}\Gamma_x}$$

dove Γ_x è il coefficiente di riflessione del carico (nel caso della TSD, dove si ha lo short S), e Γ_{M1} è il coefficiente di riflessione misurato. Andiamo avanti con i calcoli:

$$\begin{aligned} \Gamma_{M1} &= \frac{e_{00} - e_{11}e_{00}\Gamma_x + e_{10}e_{01}\Gamma_x}{1 - e_{11}\Gamma_x} = \\ &= \frac{\frac{e_{00}}{e_{10}} - \frac{e_{00}e_{11} - e_{10}e_{01}}{e_{10}}\Gamma_x}{\frac{1}{e_{10}} - \frac{e_{11}}{e_{10}}\Gamma_x} \end{aligned}$$

Dobbiamo usare, al fine di comprendere correttamente i conti, ancora una volta la trasformazione tra matrici di trasmissione e matrici scattering; come scritto in precedenza, si ha l'eguaglianza:

$$\underline{T}_A = p \begin{bmatrix} \frac{k}{p}a & b \\ \frac{k}{p} & 1 \end{bmatrix} = \frac{1}{S_{21}} \begin{bmatrix} S_{11}S_{22} - S_{21}S_{12} & S_{11} \\ -S_{22} & 1 \end{bmatrix}$$

Per semplice ispezione, è possibile dire banalmente che:

$$S_{21} = \frac{1}{p} = e_{10}$$

$$S_{11} = b = e_{00}$$

$$S_{22} = -\frac{k}{p} = e_{11}$$

$$\Delta S = -\frac{k}{p}a = e_{00}e_{11} - e_{01}e_{10}$$

Come mai ciò è interessante? Beh, perchè, come detto in precedenza, gli error box sono matrici pseudo-scattering: il formalismo è lo stesso, ma mancano di significato fisico (non ha senso fare le prove sulle perdite, sulla passività, sulla reciprocità e così via), si può dire che:

$$\Gamma_{M1} = \frac{b + \frac{k}{p}a\Gamma_x}{1 + \frac{k}{p}\Gamma_x}$$

Un ultimo “gioco”, o meglio un ultimo side-effect positivo di ciò che stiamo dicendo: si ricordi che, come appena visto:

$$S_{11} = b$$

Abbiamo preso alla leggera la cosa, ma questo fatto è molto interessante: il parametro b , che si ricava matematicamente senza bisogno di fare la misura di nessun elemento, coincide con la direttività dell'accoppiatore direzionale! Allo stesso modo, $\frac{k}{p}$ è il source match del sistema, S_{22} .

Terminato il discorso partito dalla matrice $\underline{\underline{R}}_M$, si deve fare un discorso duale con una matrice “duale”, $\underline{\underline{R}}_N$: essa è definibile come:

$$\underline{\underline{R}}_N \triangleq \underline{\underline{T}}_{MT}^{-1} \underline{\underline{T}}_{ML}$$

Effettuando gli stessi calcoli, essendo le due ancora una volta le matrici rappresentanti le misure dei due porte, si ha:

$$\underline{\underline{R}}_N = \left(\underline{\underline{T}}_{AT} \underline{\underline{T}}_{TB} \underline{\underline{T}}_{TB}^{-1} \right)^{-1} \underline{\underline{T}}_{AT} \underline{\underline{T}}_{LB} \underline{\underline{T}}_{LB}^{-1} = \underline{\underline{T}}_{BT} \underline{\underline{T}}_{TB}^{-1} \underline{\underline{T}}_{TB}^{-1}$$

Questo significa che, per la matrice $\underline{\underline{R}}_N$, $\underline{\underline{T}}_{TB}$ è la matrice degli autovettori.

Invertendo l'ordine del prodotto, si può vedere facilmente che le matrici $\underline{\underline{R}}_N$ e $\underline{\underline{R}}_M$ sono matrici **simili**: esse hanno gli autovalori coincidenti. Questo fatto non comporta tuttavia che gli autovettori siano coincidenti, cosa in generale non vera. Considerando questa volta f e g i coefficienti che si possono calcolare a partire dalla soluzione del problema agli autovalori, si consideri la matrice $\underline{\underline{T}}_{TB}$, diagonalizzante $\underline{\underline{T}}_{TL}^{-1} \underline{\underline{T}}_{TL}$, come:

$$\underline{\underline{T}}_{TB} = w \begin{bmatrix} 1 & \frac{u}{w} \\ f & \frac{u}{w}g \end{bmatrix} \triangleq w \underline{\underline{X}}_{TB}$$

Questo tipo di scrittura, come prima, è permesso, a patto di avere solo due gradi di libertà, che ora sono u e w . Dualmente a prima, ragionando dunque alla stessa maniera, si può ricavare una relazione concernente un secondo coefficiente di riflessione misurato:

$$\Gamma_{M2} = \frac{f + \frac{u}{w}g\Gamma_x}{1 + \frac{u}{w}\Gamma_x}$$

Prima di passare alle varie casistiche di taratura, è necessario introdurre un'ultima equazione matriciale: quella di de-embedding. Per fare ciò è necessario applicare la solita equazione matriciale, considerando però di avere, come oggetto tra le due matrici \underline{T}_A e \underline{T}_B^{-1} , il DUT, ossia il Device Under Test. Si avrà dunque a questo punto che la matrice delle misure effettuate sul DUT, \underline{T}_M , avrà forma del tipo:

$$\underline{T}_M = \underline{T}_A \underline{T}_{DUT} \underline{T}_B^{-1}$$

Introducendo le definizioni di matrici ricavate mediante i nostri calcoli, si può ottenere, invertendo:

$$\underline{T}_{DUT} = \underline{T}_A^{-1} \underline{T}_M \underline{T}_B = \frac{1}{p} \underline{X}_A^{-1} \underline{T}_M w \underline{X}_B \triangleq \alpha \underline{X}_A^{-1} \underline{T}_M \underline{X}_B$$

dove, ovviamente:

$$\alpha \triangleq \frac{w}{p}$$

Questa è un'equazione di de-embedding generale. È fondamentale, al fine di capire queste cose, imparare a usare questo metodo in maniera versatile: sapere come ricavare altre relazioni, a partire da misure note o dati noti. Considerando ancora una volta \underline{T}_m come la matrice delle misure del DUT RAW, si può dire che:

$$\underline{T}_M = \underline{T}_A \underline{T}_{DUT} \underline{T}_B^{-1}$$

ma, come detto prima, si può scrivere che:

$$= \underline{X}_A \underline{T}_{DUT} \frac{1}{p} \underline{T}_B^{-1} \triangleq \underline{X}_A \underline{T}_{DUT} \tilde{T}_B^{-1}$$

e come posso determinare questa nuova matrice, \tilde{T}_B^{-1} ? Niente di più semplice: si immagini per esempio di misurare il thru (scrivo direttamente la nuova forma):

$$\underline{T}_{MT} = \underline{X}_A \underline{T}_T \tilde{T}_B^{-1}$$

ma si ricordi che il thru è quel doppio bipolo di cui sono noti tutti i parametri, dunque tutta $\underline{\underline{T}}_T$! Allo stesso modo ovviamente abbiamo accesso alle misure del thru. Possiamo scrivere:

$$\underline{\underline{\tilde{T}}}_B^{-1} = \underline{\underline{T}}_M^{-1} \underline{\underline{X}}_A \underline{\underline{T}}_T$$

sostituendo:

$$\underline{\underline{T}}_{DUT} = \underline{\underline{X}}_A^{-1} \underline{\underline{T}}_M \underline{\underline{T}}_M^{-1} \underline{\underline{X}}_A \underline{\underline{T}}_T$$

dove tutte queste matrici sono note.

Riassumiamo brevemente: lo strumento non tarato introduce errori sistematici enormi, e il nostro obiettivo è fare in modo che questi si vadano a compensare. Per introdurre questa compensazione è dunque necessario introdurre un'operazione di de-embedding, che si può effettuare conoscendo un certo insieme di matrici, dipendenti da un certo insieme di parametri: $a, b, \frac{k}{p}, f, g, \frac{u}{w}, \alpha$. Come detto, per quanto riguarda i parametri a, b, f, g , è possibile semplicemente risolvere il problema agli autovettori, associato alle due matrici ottenute dal prodotto delle due matrici di trasmissione. Per fare de-embedding, tuttavia, è necessario conoscere anche gli altri tre parametri, ossia i tre rapporti: solo così è possibile avere il set di equazioni di de-embedding completo, e ottenere delle misure private di errori sistematici del DUT.

Fatta questa introduzione, è possibile passare al problema vero e proprio della taratura, ossia alla determinazione delle variabili incognite, a partire da quanto ci è noto.

TSD

A questo punto della trattazione, possiamo come detto iniziare a particularizzare: nel caso della TSD, come appena accennato, il carico è noto, ed è uno short, il cui coefficiente di riflessione sarà il nostro Γ_x , che, essendo di uno short, verrà in questo caso chiamato Γ_S . Si è appena detto che:

$$\Gamma_{M1} = \frac{b + \frac{k}{p}a\Gamma_S}{1 + \frac{k}{p}\Gamma_S}$$

$$\Gamma_{M2} = \frac{f + \frac{u}{w}g\Gamma_S}{1 + \frac{u}{w}\Gamma_S}$$

Ciò permette di ricavare, con una certa banalità, i rapporti $\frac{k}{p}$ e $\frac{u}{w}$. Manca ancora un parametro, ossia α , necessario per fare il de-embedding; per far ciò, si consideri per ipotesi di misurare il thru, applicare l'equazione di taratura,

e avere thru con lunghezza nulla (coincidente con matrice di trasmissione unitaria); si avrà:

$$\underline{\underline{T}}_{\text{MT}} = \underline{\underline{T}}_{\text{A}} \underline{\underline{T}}_{\text{T}} \underline{\underline{T}}_{\text{B}}^{-1} = \underline{\underline{T}}_{\text{A}} \underline{\underline{T}}_{\text{B}}^{-1}$$

Calcoliamo a questo punto questa matrice: ciò richiede il calcolo di una matrice inversa. Si ricorda che, per calcolare l'inversa di una matrice, è necessario:

- calcolare la matrice dei **determinanti dei minori**: dato un elemento T_{ij} , calcolare il determinante della matrice ricavata dalla matrice in questione senza le righe i e j , dunque, se $i + j$ è dispari cambiare il segno, altrimenti mantenerlo;
- farne la trasposta;
- dividere per il determinante della matrice di partenza.

Nel caso di una matrice due per due $\underline{\underline{A}}$:

$$\underline{\underline{A}} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow \underline{\underline{A}}^{-1} = \frac{1}{\det \{ \underline{\underline{A}} \}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

dunque:

$$\underline{\underline{T}}_{\text{MT}} = \frac{p}{w} \begin{bmatrix} \frac{k}{p}a & b \\ \frac{k}{p} & 1 \end{bmatrix} \begin{bmatrix} \frac{u}{w}g & -\frac{u}{w} \\ -f & 1 \end{bmatrix}$$

Dunque, si può vedere che:

$$\begin{aligned} T_{\text{MT},22} &= \frac{T_{\text{A},21}T_{\text{B},12} - T_{\text{A},22}T_{\text{B},11}}{T_{\text{B},11}T_{\text{B},22} - T_{\text{B},21}T_{\text{B},12}} = \frac{ku - pw}{wug - wuf} = \\ &= \frac{k\frac{u}{w} + p}{ug - uf} = \frac{\frac{k}{p}\frac{p}{w}u + p}{ug - uf} = \frac{\frac{k}{p}\frac{p}{w}\frac{u}{w} + \frac{p}{w}}{\frac{u}{w}(g - f)} = \\ &= \frac{\alpha^{-1}\frac{k}{p}\frac{u}{w} + \alpha^{-1}}{\frac{u}{w}(g - f)} \end{aligned}$$

Invertendo questa equazione, dove tutti i termini sono noti, è possibile calcolare il valore di α .

TRL

A differenza del caso della TSD, si ha una complicazione in più: ora invece che uno short si ha un reflect, ossia un qualcosa di **non noto**: Γ_x dunque in questo caso non sarà più una incognita, e dovremo quindi inventarci qualcosa di nuovo. Osserviamo ancora una volta le due equazioni di partenza:

$$\Gamma_{M1} = \frac{b + \frac{k}{p}a\Gamma_S}{1 + \frac{k}{p}\Gamma_x}$$

$$\Gamma_{M2} = \frac{f + \frac{u}{w}g\Gamma_S}{1 + \frac{u}{w}\Gamma_x}$$

Due equazioni, tre incognite: ci manca anche Γ_x . Come fare? Beh, si immagini di misurare un parametro scattering, per esempio $S_{M,11}$; possiamo applicare la seguente eguaglianza:

$$S_{M,11} = \frac{T_{MT,12}}{T_{MT,22}} = \frac{-T_{A,11}T_{B,12} + T_{A,12}T_{B,11}}{T_{A,21}T_{B,12} - T_{A,22} - T_{B,11}} = \frac{-kau + pbw}{ku - pw} =$$

$$= \frac{b - \frac{k}{p}\frac{u}{w}a}{-\frac{k}{p}\frac{u}{w} + 1}$$

Questa è la terza equazione che si può usare per risolvere il sistema. Si tratta di un sistema non lineare, dal momento che le incognite (i vari rapporti) si moltiplicano tra loro. Come risolverlo? Beh, si parta da quest'ultima, e la si scriva dunque nella seguente maniera:

$$\frac{k}{p}\frac{u}{w}[a - S_{M,11}] = b - S_{M,11} \implies \frac{k}{p}\frac{u}{w} = \frac{a - S_{M,11}}{b - S_{M,11}}$$

Consideriamo dunque Γ_{M2} ; ribaltandola:

$$\Gamma_{M2} \left[1 + \frac{u}{w}\Gamma_x \right] = f + \frac{u}{w}g\Gamma_x$$

da qui:

$$\Gamma_x \frac{u}{w} [\Gamma_{M2} - g] = f - \Gamma_{M2}$$

dunque:

$$\frac{u}{w} = -\frac{\Gamma_{M2} - f}{\Gamma_{M2} - g} \frac{1}{\Gamma_x}$$

in maniera duale:

$$\frac{k}{p} = -\frac{1}{\Gamma_x} \frac{\Gamma_{M1} - b}{\Gamma_{M1} - a}$$

sostituendo queste ultime due nella prima:

$$\left(-\frac{1}{\Gamma_x} \frac{\Gamma_{M1} - b}{\Gamma_{M1} - a} \right) \left(-\frac{\Gamma_{M2} - f}{\Gamma_{M2} - g} \frac{1}{\Gamma_x} \right) = \frac{a - S_{M,11}}{b - S_{M,11}}$$

ribaltando e risolvendo rispetto a Γ_x :

$$\Gamma_x = \pm \sqrt{\frac{(\Gamma_{M1} - b)(\Gamma_{M2} - f)(S_{M,11} - a)}{(\Gamma_{M1} - a)(\Gamma_{M2} - g)(S_{M,11} - b)}}$$

Una volta trovato Γ_x , usando equazioni già viste, è possibile trovare tutto il resto: il problema è sostanzialmente ricondotto alla matematica della TSD, già risolta.

Discutiamo un po' i risultati ottenuti: la TRL è una tecnica di taratura che permette di trovare un coefficiente di riflessione sconosciuto conoscendo il thru, la linea (della quale ci interessa più che altro l'impedenza caratteristica), e la fase di uno standard ad alta riflessione, del quale però non siamo interessati al modulo; la fase deve essere nota con un errore di $\pm 180^\circ$. Come mai si parla di standard ad alta riflessione? Beh, semplice: se $\Gamma_x \rightarrow 0$, le espressioni di Γ_{M1} e Γ_{M2} degenerano nelle semplici direttività, portando il sistema a essere male condizionato. Dal momento che inoltre realizzare uno short è più semplice, generalmente, che adattare, allora ben venga.

Il TRL è il sistema più accurato che ci sia: esso si basa semplicemente su un thru, che è la banale connessione delle porte, su di una linea di cui dobbiamo solo conoscere l'impedenza caratteristica Z_∞ , in modo che le sezioni trasversali siano tutte uguali, quindi uno short, che non deve essere definito (se non nella fase): è sostanzialmente necessario un solo standard, e per questo, essendo molto semplice, è anche molto precisa e utilizzata.

Ovviamente, questa tecnica di taratura non ha solo vantaggi: il problema di questa taratura è sostanzialmente la banda. Si arriva, a un certo punto, ad avere $k(l_L - l_T) = 360^\circ$, per un qualche valore delle lunghezze; questo significa che la matrice $\underline{\underline{\lambda}}$ diventa unitaria, e dunque degenera. Come si potrebbe vedere in un'analisi in frequenza, su un grafico si hanno dei picchi, dovuti a questo spiacevole effetto di risonanza tra linea e thru: dopo il de-embedding, gli errori sono comunque tali da avere risultati inaccettabili, insensati. Per evitare questo problema, si diedero specifiche sulla lunghezza in gradi degli standard: da 30° a 150° , al fine di avere margini di fase elevati. Ai tempi purtroppo non si parlava di sintetizzatori, dunque si avevano anche problemi

per quanto riguarda la ripetibilità di frequenza (tentare due volte di ottenere gli stessi identici valori di frequenza era improponibile senza sintetizzatori), dunque si tenevano di questi margini. Al giorno d'oggi, coi sintetizzatori di oggi, si può ridurre il margine di fase, arrivando a usare fino a 1° o 2° di margine.

Altre cose che si fanno riguardano la multi-line TRL, dove si usano 2 o 3 linee per fare gli standard, linee tutte di lunghezza diversa, per cercare quella con il conditioning number migliore.

3.5.2 SOLR: Short, Open, Load, Reciprocal

In diverse situazioni, una per esempio quella di connettori dello stesso sesso (femmina-femmina o maschio-maschio), fare un thru non è possibile: sarebbe necessario introdurre altri connettori, che però finirebbero per disturbare la taratura, rendendola insensata.

Come si può fare? Questo problema è un problema sostanzialmente del metodo di taratura, dunque, volendo risolverlo, conviene cambiare metodo di taratura.

Sappiamo che:

$$\underline{T}_A = p \begin{bmatrix} \frac{k}{p}a & b \\ \frac{k}{p} & 1 \end{bmatrix} \quad \underline{T}_B = w \begin{bmatrix} 1 & \frac{u}{w} \\ f & \frac{u}{w}g \end{bmatrix}$$

Con le misure alla porta 1 e alla porta 2 possiamo misurare le due matrici, basti ricordare le equazioni di taratura ribaltate e gli error box:

$$\Gamma_{M1} = \frac{b + \frac{k}{p}a\Gamma_S}{1 + \frac{k}{p}\Gamma_x}$$

$$\Gamma_{M2} = \frac{f + \frac{u}{w}g\Gamma_S}{1 + \frac{u}{w}\Gamma_x}$$

Per la prima matrice si può ricavare a, b con la soluzione del problema agli autovalori, e con le misure i $\frac{k}{p}$, effettuando tre misure di carichi noti (il corto circuito, il circuito aperto, il carico adattato come carico noto *load*); per la seconda matrice si fa la stessa cosa facendo le stesse misure, trovando i parametri $f, g, \frac{u}{w}$. Abbiamo sei misure, sette parametri incogniti: ci manca α , e ora non possiamo misurare un thru in modo tale da determinarlo. D'altra parte, ci manca ancora una condizione da imporre: *reciprocal*, ossia uno standard la cui unica cosa nota è la reciprocità, la passività. Questa condizione può essere imposta a livello di matrice di trasmissione, ricordando semplicemente il fatto che la proprietà di reciprocità sulla matrice di trasmissione impone il fatto che il suo determinante sia unitario:

$$\underline{T}_M = \frac{p}{w} \underline{X}_A \underline{T}_{DUT} \underline{X}_B^{-1} = \alpha \underline{X}_A \underline{T}_{DUT} \underline{X}_B^{-1}$$

calcolando il determinante di questa matrice, rappresentante la misura, esso deve essere unitario:

$$\det \left\{ \underline{T}_M \right\} = \alpha^2 \frac{\det \left\{ \underline{X}_A \right\} \det \left\{ \underline{T}_{DUT} \right\}}{\det \left\{ \underline{X}_B \right\}}$$

invertito:

$$\det \left\{ \underline{T}_{DUT} \right\} = \alpha^2 \frac{\det \left\{ \underline{X}_B \right\} \det \left\{ \underline{T}_M \right\}}{\det \left\{ \underline{X}_A \right\}} = 1$$

ossia, impongo la reciprocità. Ciò permette di calcolare α come:

$$\alpha = \pm \sqrt{\frac{\det \left\{ \underline{X}_B \right\} \det \left\{ \underline{T}_M \right\}}{\det \left\{ \underline{X}_A \right\}}}$$

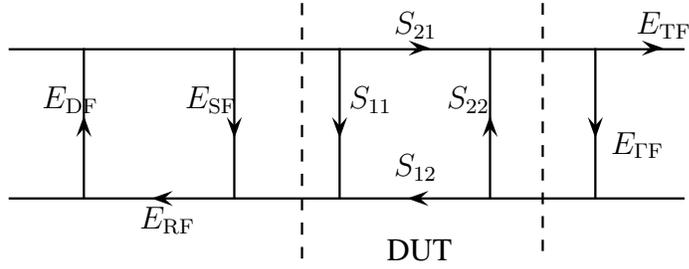
QSOLT: Quick SOLT

Non abbiamo ancora parlato di SOLT, ma una variante sul tema potrebbe essere la seguente: si tratta di una tecnica di taratura che si basa sulla misura di short, open, load alla porta 1, e di un thru, senza misurare null'altro alla porta 2, come la soluzione standard sembra suggerire.

Contando le equazioni non c'è niente di strano: mettendo tre standard alla porta 1, e quattro dati portati dal thru, si ha un sistema ben posto.

3.5.3 Approccio basato sul modello di errore forward-reverse

Finora è stato utilizzato, come modello di errore, quello basato sull'approccio "matrici di trasmissione", ossia quello a due error box; esistono altri modelli di errore, radicalmente diversi, da quello finora presentato. Questo nella fattispecie è più idoneo per un analizzatore di reti 3-sampler. Si tratta in verità di un modello doppio, dal momento che esso è costituito da due modelli distinti. Il modello usato è il seguente:



dove

b_1 è il contributo di onda riflessa alla porta 1, b_2 il contributo di onda trasmessa alla porta 2, a_{M1} la misura dell'onda incidente alla porta 1 a_1 l'onda incidente alla porta 1, e così via: il pedice M indica il fatto che l'onda sia misurata o effettivamente presente dentro al dispositivo, fisica; i pedici F e R indicano il fatto che la grandezza considerata sia del modello forward o reverse; i pedici 1 e 2 indicano semplicemente le porte considerate. Come si può vedere, questo modello ha sei parametri di errore:

- E_{FD} : errore di direttività alla porta 1;
- E_{FR} : errore di reflection tracking alla porta 1;
- E_{FS} : errore di source matching alla porta 1;
- E_{FT} : errore di riflessione alla porta 2;
- E_{FT} : errore di trasmissione alla porta 2;
- E_{FX} : errore di crosstalk tra la porta 1 e la porta 2; questo usualmente è considerato nullo, eliminando dunque un contributo e un parametro al modello (che dunque diventa a 5 parametri).

Le equazioni di questo modello possono essere scritte per ispezione.

$$b_1 = S_{11}a_1$$

Volendo misurare $S_{M,11}$, avrò:

$$S_{M,11} = \frac{b_{MR}}{a_M} = E_{FD} + \frac{E_{FS} \left(S_{11} + \frac{S_{12}S_{21}E_L}{1-S_{22}E_L} \right)}{1 - E_S \left(S_{11} + \frac{S_{12}S_{21}E_L}{1-S_{22}E_L} \right)}$$

In questo caso, il load-match, E_L , è il coefficiente di carico a quella porta: in realtà è anche lui un parametro di errore, ma non è più indipendente dal coefficiente di riflessione di sorgente, dello switch.

Questo modello, come si può vedere dalla misura del $S_{M,21}$, contiene tutti e quattro i parametri di scattering. Questo modello in totale ha 10 coefficienti: 5 di forward, 5 di reverse. L'obiettivo è sostanzialmente, date le misure dei parametri scattering $S_{M,11}, S_{M,21}, S_{M,12}, S_{M,22}$, trovare i dieci coefficienti.

Questa formulazione ha un vantaggio enorme, ossia quello dell'identificazione: per le equazioni di de-embedding la soluzione è complicatissima da ricavare. Infatti, se faccio la misura relativa a un load matchato (in cui $\Gamma_L = 0$), ottengo immediatamente che $\Gamma_{M1} = E_{FD}$, ossia identifico immediatamente il coefficiente di errore della direttività. Si noti però una cosa: questo sembrerebbe semplicemente il misurare la direttività di un accoppiatore direzionale, ma non è così, dal momento che è un **coefficiente di errore** (e le due cose coincidono se e solo se tutte le varie impedenze di riferimento sono uguali). Se poi si ha $\Gamma_S = -1$, si ha:

$$\Gamma_{M1} - E_D = \frac{-E_{FR}}{1 + E_S}$$

invece, con $\Gamma_O = +1$, si ha:

$$\Gamma_{M1} - E_D = \frac{E_{FR}}{1 + E_S}$$

Ossia, con queste semplici cose, abbiamo identificato già tre parametri, risolvendo questo sistema con queste due equazioni.

Cosa posso fare per identificare il load match? Beh, con un thru: ciò che si misura, de-embedda, è E_{FL} : la misura di S_{11} de-embeddata col thru. Dalla misura di trasmissione (S_{21}) del thru, invece, si può ricavare il E_{FT} .

Questa tecnica è ridondante: si hanno 10 incognite, e 10 equazioni, ossia 10 misure; non si può vedere la ridondanza, scrivendo in questo modo: il modello error-box permette di capirlo. Questo modello ha vantaggi: si devono misurare solo 3 onde, e non 4 (per 2, ossia sia per forward sia per reverse): a_M, b_{MT}, b_{MR} , dunque si può applicare anche al caso dell'analizzatore 3-sampler, dal momento che non siamo interessati alla quarta misura.

3.5.4 SOLT applicata a un multiporte

Ciò che si farà sarà ricavare l'equazione per un multiporte, usando una formulazione basata sulla matrice scattering: la matrice di trasmissione presenta, nel caso di match, non esiste: non si può scrivere la matrice di trasmissione di load, dal momento che avrebbe ∞ sulla colonna: non si può applicare il formalismo TRL a ciò. Data una formulazione multiporte non si è più vincolati a una particolare formulazione.

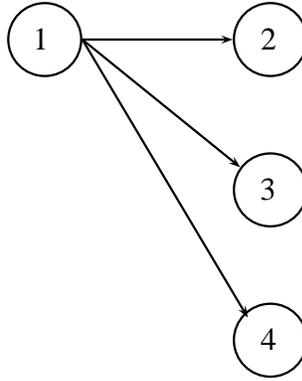
A multiporte non esistono standard: più che lo standard fully-unknown, non siamo in grado di avere nulla: gli standard si considerano a due porte. Si ha un secondo constraint: per ragioni fisiche è possibile che non si riesca a collegare una certa porta a un'altra porta, ossia che non sia possibile fare dei thru, per semplici ragioni fisiche o di posizioni delle porte su di uno strumento. È fondamentale dunque una formulazione in grado di svincolarci dal dover mettere standard in un posto piuttosto che in un altro, formulazione basata solo su standard a due porte, e che include la possibilità di avere standard reciproci, e questo tutto a multiporte.

3.6 Taratura di analizzatori di reti multiporte

Ciò che si farà sarà ricavare l'equazione per un multiporte, usando una formulazione basata sulla matrice scattering: la matrice di trasmissione presenta, nel caso di match, non esiste: non si può scrivere la matrice di trasmissione di load, dal momento che avrebbe ∞ sulla colonna: non si può applicare il formalismo TRL a ciò. Data una formulazione multiporte non si è più vincolati a una particolare formulazione.

A multiporte non esistono standard: più che lo standard fully-unknown, non siamo in grado di avere nulla: gli standard si considerano a due porte. Si ha un secondo constraint: per ragioni fisiche è possibile che non si riesca a collegare una certa porta a un'altra porta, ossia che non sia possibile fare dei thru, per semplici ragioni fisiche o di posizioni delle porte su di uno strumento. È fondamentale dunque una formulazione in grado di svincolarci dal dover mettere standard in un posto piuttosto che in un altro, formulazione basata solo su standard a due porte, e che include la possibilità di avere standard reciproci, e questo tutto a multiporte.

Questa formulazione deve essere la generalizzazione di formulazioni già esistenti: riducendo gli indici, di fatto si deve tornare alle già note formulazioni; supponendo per esempio di chiedere che il numero N di porte sia pari a 1, la formulazione deve degenerare a quella del 1-porta. Potremo sfruttare, dopo aver introdotto questa formulazione, alcuni principi concettualmente già noti: il fatto, per esempio, di poter estendere facilmente la QSOLT in un N -porte.



Introdurre i tre standard alla porta 1, dunque il thru introduce 4 equazioni a ogni porta che aggiungiamo.

3.6.1 Formulazione relativa ad analizzatori 4-sampler

La formulazione che studieremo è sostanzialmente l'estensione del modello 1-porta: nel modello a 1 porta si aveva un'equazione che fissava un legame tra le onde reali a quelle misurate; si ha dunque un set di equazioni di questo tipo:

$$\begin{cases} a_i = l_i b_{m,i} - h_i a_{m,i} \\ b_i = k_i b_{m,i} - m_i a_{m,i} \end{cases}$$

Questo è il modello hardware che permette di descrivere una porta; queste equazioni possono essere scritte in forma matriciale, considerando dunque il caso più generale, ottenendo dunque un sistema di questo tipo:

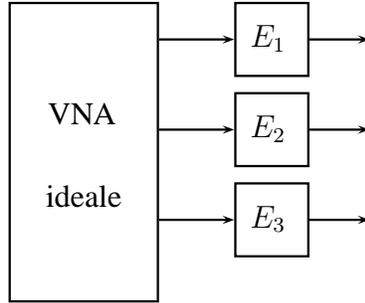
$$\begin{cases} \underline{a} = \underline{L}\underline{b}_m - \underline{H}\underline{a}_m \\ \underline{b} = \underline{K}\underline{b}_m - \underline{M}\underline{a}_m \end{cases}$$

Si noti che, finora, non si è parlato di switch correction. Se al giorno d'oggi, con gli strumenti attualmente in commercio, si misurano onde di potenza, ai tempi in cui questa formulazione era stata proposta si misuravano parametri scattering, portando dunque a risultati diversi. Il nostro obiettivo è, usando una formulazione relativamente semplice, ottenere gli stessi risultati trovati 20 anni fa. Come noto, esiste la relazione:

$$\underline{b} = \underline{S}\underline{a}$$

Dal momento che ciascuna onda nel dispositivo reale è collegata alle misure delle medesime porte e a nessun'altra, le quattro matrici \underline{L} , \underline{H} , \underline{K} , \underline{M}

saranno tutte diagonali. Questa cosa può avere un significato interessante: è come se, dato un VNA in grado di misurare idealmente tutte le uscite, si avessero degli error box a due porte per ciascuna misura:



Il VNA ideale dovrebbe misurare parametri scattering, dunque i risultati sono un po' più complicati; con le onde di potenza, tutto diventa più facile. Ricordando le definizioni di parametri scattering, si ha che, per calcolare ciascun parametro scattering, eccito una porta e chiudo tutte le altre su carico adattato, trovando una colonna per ogni porta eccitata. Si avrà dunque qualcosa di questo genere:

$$[\underline{b}' \ \underline{b}'' \ \dots \ \underline{b}^n] = \underline{\underline{S}} [\underline{a}' \ \underline{a}'' \ \dots \ \underline{a}^n]$$

definendo due matrici relative a quelle appena viste, la stessa equazione di qui in poi sarà scritta, per compattezza, come:

$$\underline{\underline{B}} = \underline{\underline{S}} \underline{\underline{A}}$$

A questo punto, $\underline{\underline{B}}$ e $\underline{\underline{A}}$ sono le matrici contenenti tutti i risultati delle misure alle varie porte. Ciò permette di scrivere il sistema finale comprendente tutte le misure/onde alle varie porte:

$$\begin{cases} \underline{\underline{B}} = \underline{\underline{KB}}_m - \underline{\underline{MA}}_m \\ \underline{\underline{A}} = \underline{\underline{LB}}_m - \underline{\underline{HA}}_m \end{cases}$$

Usando la relazione comprendente la matrice scattering, si ottiene:

$$\left(\underline{\underline{KB}}_m - \underline{\underline{MA}}_m \right) = \underline{\underline{S}} \left(\underline{\underline{LB}}_m - \underline{\underline{HA}}_m \right)$$

ciò permette di scrivere l'equazione di de-embedding:

$$\underline{\underline{S}} = \left(\underline{\underline{KB}}_m - \underline{\underline{MA}}_m \right) \left(\underline{\underline{LB}}_m \underline{\underline{HA}}_m \right)^{-1}$$

Questa è l'equazione di de-embedding per i multiporta, scritta senza switch correction, direttamente per le onde.

Recuperando l'equazione di prima, si ha, ordinando i termini:

$$\underline{\underline{KB}}_m \underline{\underline{MA}}_m - \underline{\underline{SLB}}_m + \underline{\underline{SHA}}_m = \underline{\underline{0}}$$

definendo una matrice $\underline{\underline{B}}_m = \underline{\underline{S}}_m \underline{\underline{A}}_m$, sostituendo dentro la formula e moltiplicando tutto a destra per $\underline{\underline{A}}_m^{-1}$, si ottiene:

$$\underline{\underline{KS}}_m - \underline{\underline{M}} - \underline{\underline{SLS}}_m + \underline{\underline{SH}} = \underline{\underline{0}}$$

A questo punto, vediamo se questa cosa è “downsizeabile”: proviamo a scrivere, per questo modello, l'equazione corrispondente a questa, ma solo a 1 porta. Si ottiene qualcosa di questo tipo:

$$\Gamma_{m1} = \frac{M_1}{\Gamma_1} + \frac{\Gamma_1 L_1}{K_1} - \frac{\Gamma_1 M_1}{K_1}$$

che si può scrivere semplicemente come:

$$(1 - \tilde{L}\Gamma) \Gamma_{m1} = \tilde{M} - \Gamma \tilde{H}$$

ossia

$$\Gamma_m = \frac{\tilde{M} - \tilde{H}\Gamma}{1 - \tilde{L}\Gamma}$$

Questo assomiglia molto a quanto misurato precedentemente, parlando di modello di error box a 1 porta:

$$\Gamma_m = \frac{E_D - \Delta\Gamma}{1 - E_S\Gamma}$$

Come si vede, si potrebbero trovare analogie tra i parametri dell'error box ricavati mediante le matrici di trasmissione e questo. \tilde{L} è circa uguale al source match, \tilde{H} al determinante, \tilde{M} alla direttività.

Dal momento che le misure che facciamo sono misure di rapporto, è possibile introdurre una normalizzazione in ciò che misuriamo: quello che si fa di solito, canonicamente, è normalizzare $K_1 = 1$, dunque di tutta la matrice $\underline{\underline{K}}$, la quale sarà una matrice diagonale (come le altre tre “matrici error box”), un coefficiente (il primo) sarà unitario, e tutti gli altri (della stessa matrice e delle altre matrici) saranno normalizzati rispetto a esso.

Con la nostra formulazione abbiamo dunque 4 matrici incognite, la dimensione di ciascuna delle quali dipende sostanzialmente dal numero di porte presenti; si ha che inoltre queste matrici sono diagonali, dunque ciascuna matrice introduce un numero di incognite sostanzialmente pari al numero di porte presenti nello strumento. Riassumendo, il numero di parametri sarà:

$$N_{\text{parametri}} = N_{\text{porte}} \times 4 - 1$$

il -1 deriva dal fatto che tutto è normalizzato rispetto a K_1 , dunque si ha un parametro in meno. Con un 4 porte, si hanno 15 parametri dunque.

3.6.2 Formulazione relativa ad analizzatori 3-sampler

Prima di andare avanti col problema, è necessario considerare un'altra casistica: quella per cui si ha a che fare con analizzatori di reti basati sullo schema 3-sampler. Questo schema concettualmente è, come noto, ben diverso dagli altri, dal momento che prima per ogni porta avevamo due misure, dateci dai due accoppiatori direzionali, ma negli analizzatori di reti a 3 sampler ciò non è vero: si ha un solo accoppiatore per ciascuna porta, e lo schema risulta essere diviso in due metà, ciascuna delle quali ha un solo riferimento. Prima che si rifacesse il modello di errore (quello che stiamo per presentare), al fine di avere un numero sufficiente di parametri per la taratura, si utilizzavano semplicemente tutte le possibili combinazioni a coppie delle porte: si mettevano assieme tanti modelli a 2 porte.

Rispetto al caso di prima, le porte possono essere in due condizioni, in due stati:

- stato A: stato in cui sorgente e riferimento sono collegati alla porta i -esima; in questo stato, non si ha nulla da aggiungere rispetto al caso precedente, nel senso che, per questo stato, si avrà di nuovo:

$$\begin{cases} a_i = l_i b_{m,i} - h_i a_{m,i} \\ b_i = k_i b_{m,i} - m_i a_{m,i} \end{cases}$$

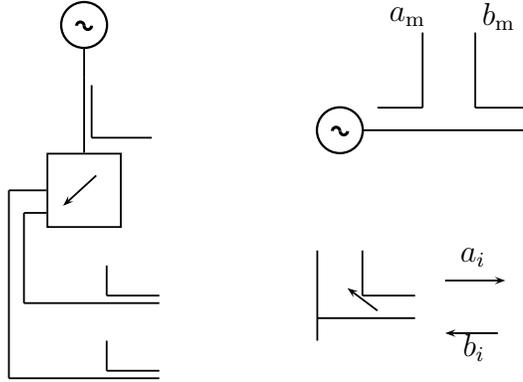
e le varie estensioni vettoriali esattamente analoghe a prima; per lo stato A, dunque, considereremo le equazioni generali in forma matriciale scritte nella seguente maniera:

$$\begin{cases} \underline{\tilde{B}} = \underline{K}\underline{\tilde{B}}_m - \underline{M}\underline{\tilde{A}}_m \\ \underline{\tilde{A}} = \underline{L}\underline{\tilde{B}}_m - \underline{H}\underline{\tilde{A}}_m \end{cases}$$

- stato B: in questo caso la i -esima porta non è collegata alla sorgente, bensì è chiusa sul carico.

Conosciamo dunque già le equazioni per lo stato A, non conosciamo quelle per lo stato B, ma sappiamo come ricavarle con semplicità: prima di tutto,

quando abbiamo ricavato gli error box per l'1-porte, l'ipotesi di base era la sua linearità, dunque siamo sicuri di aver a che fare con equazioni lineari; oltre a questo, sappiamo che lo schema sarà una cosa di questo tipo:



Dal momento che si ha, come unica variabile indipendente, l'onda misurata, si dovrà per forza avere qualcosa del tipo:

$$\begin{cases} a_i = g_i \hat{b}_{m,i} \\ b_i = f_i \hat{b}_{m,i} \end{cases}$$

Queste equazioni sono ovviamente generalizzabili nelle seguenti:

$$\begin{cases} \underline{\hat{A}} = \underline{G} \underline{\hat{B}}_m \\ \underline{\hat{B}} = \underline{F} \underline{\hat{B}}_m \end{cases}$$

A questo punto, cosa abbiamo? Sostanzialmente, due modelli che però dovrebbero far parte di un singolo modello: esso dovrebbe essere rappresentato mediante singole matrici $\underline{\hat{A}}$ e $\underline{\hat{B}}$. Ciò che abbiamo ora tuttavia sono due coppie di matrici come variabili dipendenti, e queste sono relative ciascuna dallo stato, e hanno caratteristiche diverse. Dobbiamo legare dunque i due modelli, al fine di ottenere un solo modello. I valori di a_i e b_i non devono dipendere dal modello che dobbiamo utilizzare: dobbiamo solamente avere delle matrici in grado di rappresentare tutto ciò come già dovrebbe essere. Le onde al piano di riferimento devono essere indipendenti dalla situazione. L'obiettivo è trovare dunque un'equazione con la matrice scattering (funzione di a_i e b_i per ogni porta) che tenga conto di tutto ciò.

Osserviamo: se mettiamo la sorgente a una certa porta, per esempio la porta 1, questo significa che le altre porte sono nello stato B (dal momento che esse sono caricate su di un carico, non sulla sorgente); questo significa che non si ha, per ciascuna di queste porte, la $a_{m,i}$, $i \neq 1$: su ciascuna delle porte in cui non si ha la sorgente, non si ha nulla in grado di misurare l'onda incidente, dunque si hanno solo le b_i (il modello è a 3 sampler). Negli stati

A ho solamente la $\underline{\hat{A}}_m$ e $\underline{\hat{B}}_m$, le quali hanno solo gli elementi sulla diagonale; dualmente, lo stato \underline{B} non può misurare l'onda di sorgente, dunque le sue matrici avranno solo elementi al di fuori della diagonale: le $\underline{\hat{B}}_m$ e $\underline{\hat{A}}_m$ saranno matrici senza elementi sulla diagonale.

Nessuno ci impedisce dunque di fare lo stesso ragionamento delle onde misurate sulle onde vere: considerarne una parte solo sulla diagonale, una parte solo fuori dalla diagonale, e considerare il modello finale come somma delle due matrici:

$$\underline{\underline{A}} = \underline{\hat{A}} + \underline{\tilde{A}}$$

$$\underline{\underline{B}} = \underline{\hat{B}} + \underline{\tilde{B}}$$

Queste sono le matrici che possiamo usare per questo modello “generalizzato”, valido per gli analizzatori 3-sampler: è da questa idea, che permette di congiungere i due stati, che è possibile raggiungere il modello finale. Riprendendo le definizioni prima viste, avremo banalmente:

$$\underline{\underline{A}} = \underline{\hat{A}} + \underline{\tilde{A}} = \underline{\underline{L}}\underline{\tilde{B}}_m - \underline{\underline{H}}\underline{\tilde{A}}_m + \underline{\underline{G}}\underline{\hat{B}}_m$$

$$\underline{\underline{B}} = \underline{\hat{B}} + \underline{\tilde{B}} = \underline{\underline{K}}\underline{\tilde{B}}_m - \underline{\underline{M}}\underline{\tilde{A}}_m + \underline{\underline{F}}\underline{\hat{B}}_m$$

Fatto ciò, siamo interessati a ricavare l'equazione di de-embedding dei parametri scattering, semplicemente sostituendo le espressioni appena viste in un'equazione del tipo:

$$\underline{\underline{B}} = \underline{\underline{S}}\underline{\underline{A}}$$

Facendo ciò, si ottiene un'equazione in grado di porre in relazione le altre due equazioni, semplicemente come:

$$-\underline{\underline{S}}\underline{\underline{G}}\underline{\hat{B}}_m + \underline{\underline{F}}\underline{\hat{B}}_m - \underline{\underline{S}}\underline{\underline{L}}\underline{\tilde{B}}_m + \underline{\underline{K}}\underline{\tilde{B}}_m + \underline{\underline{S}}\underline{\underline{H}}\underline{\tilde{A}}_m - \underline{\underline{M}}\underline{\tilde{A}}_m = \underline{\underline{0}}$$

si noti che, se $\underline{\underline{B}} = \underline{\underline{0}}$, questo modello si riconduce a quello visto per il 4-sampler. In questo caso si hanno, per questa equazione, 6 matrici: $\underline{\underline{G}}, \underline{\underline{F}}, \underline{\underline{L}}, \underline{\underline{K}}, \underline{\underline{H}}, \underline{\underline{M}}$; esse saranno tutte di nuovo diagonali, dunque per ciascuna di esse vi sarà un numero di parametri pari al numero di porte presenti, ed esse sono 6. Ora questo modello ha:

$$N_{\text{parametri}} = N_{\text{porte}} \times 6 - 1$$

Questa formula è tendenzialmente giusta, ma non perfetta: se considerassimo per esempio un VNA a 2 porte, si avrebbe:

$$2 \times 6 - 1 = 11$$

quando in realtà, come già visto in precedenza, di parametri ne servono in realtà solamente 10. In realtà non si ha nulla di sbagliato, dal momento che ci manca un'informazione apparentemente invisibile: il caso a 2 porte è un caso degenere per questo modello. Si vede che le due equazioni in $\underline{\underline{F}}$ e in $\underline{\underline{G}}$ si disaccoppiano, e dunque si finisce per scrivere esattamente ciò che si scriveva con il modello meno generale precedentemente introdotto: si vede che le parti dipendenti da $\underline{\underline{H}}, \underline{\underline{K}}, \underline{\underline{L}}, \underline{\underline{M}}$ sono accoppiate a $\underline{\underline{F}}, \underline{\underline{G}}$ sia alla porta 1, sia alla porta 2, ma non ci sono parti in cui $\underline{\underline{F}}$ e $\underline{\underline{G}}$ della porta 1 son accoppiate anche alle $\underline{\underline{F}}$ e $\underline{\underline{G}}$ alla porta 2.

3.6.3 Calcolo degli error coefficients

A questo punto è possibile iniziare a capire come calcolare i coefficienti di errore per le nostre formulazioni. Quello che abbiamo finora tirato fuori è l'equazione di taratura:

$$-\underline{\underline{S}}\underline{\underline{G}}\underline{\underline{\hat{B}}}_m + \underline{\underline{F}}\underline{\underline{\hat{B}}}_m - \underline{\underline{S}}\underline{\underline{L}}\underline{\underline{\tilde{B}}}_m + \underline{\underline{K}}\underline{\underline{\tilde{B}}}_m + \underline{\underline{S}}\underline{\underline{H}}\underline{\underline{\tilde{A}}}_m - \underline{\underline{M}}\underline{\underline{\tilde{A}}}_m = \underline{\underline{0}}$$

A questo punto, alcune osservazioni:

- è basata sui parametri scattering (mentre quella con la quale siamo familiari aveva le matrici di trasmissione);
- è definita per qualsiasi standard;
- può essere per trovare, dato un insieme di standard noti, i parametri $\underline{\underline{H}}, \underline{\underline{L}}, \underline{\underline{M}}, \underline{\underline{K}}, \underline{\underline{F}}, \underline{\underline{G}}$, ossia i “coefficienti di errore”;
- può essere usata, data la calibrazione, ossia una volta calcolati i coefficienti di errore, per determinare $\underline{\underline{S}}$.

A questo punto, come si fa ad andare avanti? L'idea fondamentale è quella di prendere l'equazione di calibratura e scriverla in forma scalare. Come mai un passaggio di questo tipo? Beh, come già detto nell'introduzione, all'inizio avevamo diverse necessità, una delle quali era quello di utilizzare degli standard a **due porte**: non esistono standard a N porte! Ricordando poi che l'analizzatore di reti è suddiviso in due parti, le due porte possono essere tutte e due a sinistra, tutte e due a destra, una a sinistra e una a destra, o anche a una porta. Scalarizzando le equazioni, non abbiamo più problemi a dover scrivere per intero la matrice scattering di standard che probabilmente

non abbiamo neanche a disposizione fisicamente, bensì dovremo semplicemente scrivere un insieme di equazioni (2, 4, 6) a seconda di come misuriamo lo standard a 1 o 2 porte.

Si consideri un esempio: si immagini di mettere un due porte standard (un thru) tra due porte appartenenti alla metà di destra dello strumento; si avrà lo stato A da una parte, lo stato B dall'altro. In questa situazione, l'oggetto dà origine a 4 equazioni: 2 per direzione. Si ha qualcosa di analogo alla SOLT. Si può infatti fare, come misure:

$$\frac{b_1}{a_1} \quad \frac{b_2}{a_1} \quad \frac{b_1}{a_2} \quad \frac{b_2}{a_2}$$

Quando siamo in stato A-B, si hanno 4 equazioni di questo tipo (3-sampler). In stato A-A si hanno quattro equazioni del tipo "quattro sampler". Quando si usa una porta di destra con quella di sinistra, si hanno 8 equazioni: 4 nello stato A, 4 nello stato A-B (si deve fare A-B e B-A).

Si parla di dynamic calibration: essa è una calibrazione che prescinde da una fissa sequenza di standard. Essendo capaci a programmare, basandoci su questo principio, potremmo scrivere dunque un algoritmo che fornisce una sequenza di campioni non fissa come nel caso della TRL, bensì che dipende da constraints che introduciamo: quale porta può essere collegata con quale altra, e con quali standard. Inoltre, con questa idea, è possibile usare le tecniche di taratura già note a 2 porte, usando standard incogniti: è sostanzialmente come ragionare a 2 porte più volte. Quando si ha a che fare con stati A-A, si possono usare tutte le nozioni note dalle tarature a 2 porte 4-sampler (per esempio, la TRL); nel caso in cui si ha a che fare con altri stati, quali A-B o B-A, è sufficiente fare quindi le misure di $\underline{\underline{F}}$ e $\underline{\underline{G}}$ per ogni porta mettendo lo standard B e facendo misura extra:

$$\underline{\underline{F}} = \underline{\underline{\hat{A}\hat{B}}}^{-1}$$

$$\underline{\underline{G}} = \underline{\underline{\hat{B}\hat{B}}}^{-1}$$

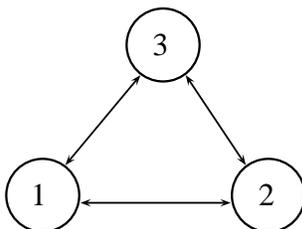
Si possono mischiare self-calibration con tarature globali, e risolvere dunque in diverse maniere diversi problemi.

Alcune note sull'indipendenza lineare delle equazioni introdotte dagli standard

Esistono dei criteri empirici che permettono di dire quante equazioni indipendenti danno un certo insieme di standard.

Una prima domanda potrebbe essere: quante equazioni linearmente indipendenti danno N standard a 1 porta? Si immagini di avere 10 load, 5 short, 20 open; di equazioni indipendenti, dato un N porte, se ne hanno $3 \times N$: le equazioni che scriviamo a 1 porta possono al più determinare i parametri di un error-box a 1 porta, dunque indipendentemente da quanti se ne mettono, il risultato finale sarà sempre lo stesso.

Un discorso più interessante potrebbe essere quello riguardante i 2-porte: capire quante equazioni linearmente indipendenti sono fornite da uno standard a 2 porte, ossia da un thru. Gli standard a 2 porte possono essere collegati in svariati modi; si consideri per esempio il caso di un 3-porte:



Questa struttura, detta **thru loop**, sembrerebbe portare a $4 + 4 + 4 - 1$ equazioni, dunque a 11 equazioni. Queste 12 equazioni introdotte, sono sul serio tutte linearmente indipendenti? Beh, da 12 iniziamo a toglierne 1 per la normalizzazione, arrivando già a 11; in realtà di queste 11 solamente 10 sono linearmente indipendenti, ma dimostrarlo in maniera formale è molto complicato.

Ciò che possiamo fare, al fine di fissare questo concetto, è dimostrarlo in maniera meno formale ma probabilmente più efficace. Partendo dal 2-porte, si immagini di avere un thru, ossia un doppio bipolo completamente determinato, il quale tuttavia non ha proprietà di reciprocità: la sua matrice scattering non ha particolari proprietà. Questo fatto porterebbe (supponendo di avere uno standard di questo genere e di averlo caratterizzato correttamente) a ottenere, quando si collega la porta 1 del thru alla porta 1 dell'analizzatore e la porta 2 all'altra, e la porta 1 del thru alla porta 2 dell'analizzatore e la 2 alla 1, ad avere 4 equazioni in un verso, e 4 equazioni nell'altro: un totale di 8 equazioni. Di queste, non tutte sono linearmente indipendenti: come si era visto quando si parlava di modelli basati sull'uso delle matrici di trasmissione, si è visto che la definizione del problema agli autovettori portava all'introduzione di due gradi di libertà, che riducono di fatto il rango del sistema di due ordini: il rango del sistema sarà pari a 6 e non a 7 o 8 (abbiamo 8 equazioni, ce ne servono 7 per avere un sistema completamente determinato, ma ne abbiamo 6). Per un 3-porte, si aggiungono quattro incognite, ma il thru alla terza porta aggiunge altre quattro equazioni che saranno

completamente indipendenti, facendoci arrivare così a 10 equazioni; questa è una dimostrazione semplificata del concetto. Il thru-loop, che si può fare senza problemi quando i connettori sono senza sesso (per esempio gli APC7), danno dunque origine a 10 equazioni linearmente indipendenti.

Ci manca un'equazione: sia per il 2-porte, sia per il 3-porte, ci manca un'idea, anche supponendo di avere a disposizione un thru non degenero: il thru non simmetrico, come detto, deve essere ben caratterizzato, e dunque si avrebbero problemi di ripetibilità.

Dove posso prendere l'equazione mancante? Beh, per rispondere alla domanda, si potrebbe dare il seguente indizio: collegando uno short a una delle porte, l'equazione che ne scaturisce non è linearmente indipendente; collegando un **offset short** (ossia uno short con un tratto di linea di trasmissione; esso può essere fatto in guida d'onda con grande precisione e poche perdite) o un load, l'equazione è linearmente indipendente dalle altre. L'idea è: quello che manca alle nostre equazioni è l'imposizione di una certa impedenza di riferimento: noi dobbiamo misurare dei parametri scattering, i quali esistono solo se definiti rispetto a una certa impedenza di riferimento; se si mettesse uno short, qualsiasi impedenza di riferimento andrebbe bene, dunque essa continuerebbe a rimanere un grado di libertà, e la cosa non è buona, dal momento che si continuerebbe a dover normalizzare rispetto a un valore scelto da noi arbitrariamente le misure, ottenendo però misure non calibrate. Tutte le tarature viste, se ci si pensa, hanno un qualcosa che fissi l'impedenza caratteristica: LRM ha il Load, TRL la Line, SOLR il Load. Nel caso multiporte, non siamo interessati alla sequenza o al numero di standard, sempre supponendo di avere dei thru a sufficienza, ma è fondamentale avere uno standard che fissi l'impedenza di riferimento.

Dato questo accorgimento, il thru-loop è probabilmente la migliore tecnica di calibrazione esistente: essa usa il minimo numero di standard possibile, dunque l'incertezza sarà tendenzialmente minore delle altre; questo non è stato ancora dimostrato, ma c'è fiducia in tale direzione.

Una nota aggiuntiva sugli offset short: usando 3 offset short, è possibile ottenere una delle tarature più accurate possibili nel caso in cui si abbia solo una porta. Il problema della cosa è però il fatto che l'offset short ha una linea di offset, dunque, come nel caso della TRL, essa tende a risuonare, per certi valori di lunghezza, con lo short.

3.6.4 Cause di incertezza

Analizziamo dunque le varie cause.

Repeatability

I problemi legati alla ripetibilità della misura possono riguardare varie cose. Una causa di sicuro potrebbe riguardare le caratteristiche meccaniche dei connettori. Se per esempio si avesse ± 1 mils, ossia $25 \mu\text{m}$, si ha qualcosa del tipo (supponendo 10 GHz):

$$\beta l = \frac{2\pi f}{c} \frac{180}{\pi} = 0,3^\circ$$

Questo significa che solo errori sul pin portano a quasi mezzo grado di differenza. Nei wafer probe si potrebbe avere anche di peggio: la spaziatura sarà anche minore, però ci sono problemi per esempio nella definizione dei materiali: il contact e la resistenza di contatto dipendono infatti da che materiale e dalla deposizione sul substrato; le probes possono essere fatte in copper-berillium; una volta si avevano con un circuito coplanare in alumina montato al contrario. Ciò che si usa di più al giorno d'oggi sono dei microcoassiali, in cui il ground è rigido.

Il caso dei sistemi multiporte ovviamente è ancora peggiore: si hanno molte più probe, dunque gli errori dovuti ai disallineamenti si moltiplicano ancora.

Modello di ripetibilità

La ripetibilità si può modellare mediante un doppio bipolo random, ossia i cui parametri (nel nostro caso, scattering), sono variabili casuali. Si avrà dunque qualcosa di questo tipo:

$$\underline{\underline{R}} = \begin{bmatrix} \delta_R & 1 + \delta_T \\ 1 + \delta_T & \delta_R \end{bmatrix}$$

Siccome tutto è “piccolo”, possiamo pensare a ciò così. Le due δ_R e δ_T sono variabili casuali.

Questa matrice può essere utilizzata per tenere conto del fatto che i coefficienti dell'error box variano in maniera aleatoria.

Standard accuracy

Oltre al problema della ripetibilità, un problema random che ci affligge è quello dell'accuratezza con la quale misuriamo gli standard. Ciò che si fa di solito al fine di caratterizzare uno standard è utilizzare una simulazione elettromagnetica, per esempio mediante metodo FEM. In microstriscia, la cosa è più problematica: come noto, quando si ha a che fare con degli standard da caratterizzare, una delle tecniche più utilizzate è basata sul metodo

FEM, ossia su di una discretizzazione dello spazio, per poi passare a un metodo numerico la soluzione del problema elettromagnetico. Se lo standard è su microstriscia, è necessario rappresentare tutte le varie dimensioni e introdurre una mesh per ciascuna di esse; si può avere a che fare con un dielettrico spesso per esempio $750 \mu\text{m}$, e con lo spessore della striscia di $20 \mu\text{m}$: se da un lato si sa che esistono algoritmi che regolano automaticamente il passo, dall'altro non si può fare più di tanto, e se si chiede che il mesh sia troppo stretto si rischia di non ottenere convergenza dell'algoritmo. Esistono poi altri problemi, come la rugosità superficiale, o la ε_r (la quale può essere stimata con misure a basse frequenze di capacità, o misurando la frequenza di risonanza introducendo un blocco di materiale noto).

Una tattica spesso utilizzata è quella di effettuare, per le resistenze presenti, un trimming laser: si brucia una parte del carico fino a ottenere esattamente il valore di resistenza desiderato. Questo purtroppo non ci permette, per quanto riguarda standard costituiti da resistori e linee, di migliorare arbitrariamente le cose: ciò che potrebbe essere problematico potrebbe essere anche la linea e non il carico.

Quando si ha la possibilità di fare le cose per bene, la cosa migliore è realizzare gli standard su substrati di quarzo, fatti bene, con misure in DC della resistenza, normalizzando tutto rispetto a essa: in questa maniera, si trovano delle buone tarature.

In questo caso, ossia l'elenco delle specifiche relative al 8510, si utilizza un modello sostanzialmente simile a quello inizialmente proposto: si ha un error box definito da parametri casuali; si può vedere la presenza di un certo numero di parametri: residual directivity, source match, e così via. Si può vedere che la direttività residua, per esempio, è di 48 dB. Si noti che purtroppo neanche un grafico ha andamento lineare: è tutto non lineare.

Controllo di accuratezza 1-porta

Si vuole a questo punto proporre un calcolo finalizzato al controllo di accuratezza: verificare il $\Delta\Gamma$. Si ha che:

$$\begin{cases} b_{m1} = E_D a_{m1} + E_R b_1 \\ a_1 = a_{m1} + E_S b_1 \end{cases}$$

Questo permette di calcolare:

$$\Gamma_{m1} = \frac{b_{m1}}{a_{m1}} = E_D + \frac{E_R \Gamma_x}{1 - E_S \Gamma_x}$$

Posso definire il $\Delta\Gamma$ come:

$$\Delta\Gamma \triangleq \Gamma_{m1} - \Gamma_x = \frac{E_D - E_S E_D \Gamma_x + (E_R - 1)\Gamma_x + E_S \Gamma_x^2}{1 - E_S \Gamma_x}$$

Di questo, sono interessanti sostanzialmente due valori: quello per cui $\Gamma_x \sim 0$, ossia quando siamo in una condizione prossima a quella di adattamento, e quello per cui $\Gamma_x \sim 1$, ossia quando siamo in una condizione prossima alla riflessione di tutta l'onda. Per $\Gamma_x \sim 0$, sviluppando secondo McLaurin l'espressione, ricordando che:

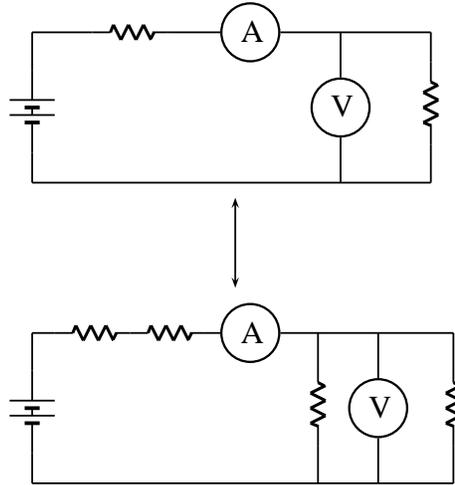
$$\frac{1}{1-x} \sim 1+x$$

si ottiene:

$$\Delta\Gamma \sim E_D + (E_R - 1)\Gamma_x + E_S \Gamma_x^2$$

Questa è un'equazione parabolica, che dipende dal valore di Γ_x . Come mai qualcosa del genere?

Proviamo a fare un passo indietro, e a collegare i concetti legati all'analizzatore di reti con quelli appartenenti a un qualcosa di più semplice: l'analizzatore di reti, in un certo senso, è una versione a microonde di un sistema voltamperometrico: la misura di coefficienti scattering invece che resistenze.



Esattamente come in un sistema voltamperometrico, si avranno non idealità: la batteria avrà una resistenza equivalente di uscita, l'amperometro avrà una resistenza di consumo, e anche l'amperometro: ciò va a entrare, nel network analyzer, nell'error box. R_x è la resistenza incognita.

Si hanno, come si vede dal secondo schema, delle resistenze parassite, che sono state "spostate" in modo da rendere più banale il ragionamento. I

parametri parassiti sono dei parametri residui, nel nostro sistema, in seguito alla taratura. Se la resistenza parallela è prossima a 0, il voltmetro misura una tensione bassissima, dal momento che su di essa cade poca tensione, e dunque si avranno enormi incertezze di misure; una qualunque inaccuratezza sulla stima dei parametri di consumo si tradurrà immediatamente in un'incertezza molto più elevata sulla misura di tensione. Caso duale è quello per cui la resistenza serie dell'amperometro è molto grande: ho poca corrente, dunque l'amperometro misura poca corrente, e quindi il parametro di consumo di nuovo porta a enormi incertezze. Quando si ha il 50Ω , si è nella situazione ottimale: la tensione misurata sarà buona (non il meglio, sarebbe meglio avere un open ma ciò porterebbe problemi dall'altra parte), e la corrente idem come appena detto. Questa è una situazione abbastanza buona per la stima della resistenza, o del Γ nella situazione dell'analizzatore di reti.

Il fatto di sviluppare a 0 o a 1 la funzione di $\Delta\Gamma$ porta sostanzialmente a due situazioni duali, in qualche modo riconducibili al metodo voltamperometrico (per quanto qua si richiede conoscenza dei numeri complessi): di fatto, nelle due situazioni, si avranno due problemi di tipo diverso: qua si avranno modulo e fase, e nelle due situazioni si avrà:

- nel caso di $|\Gamma| \sim 0$, il modulo sarà molto preciso, ma la fase potrebbe variare enormemente: avere il modulo del coefficiente di riflessione prossimo a 0 significa sostanzialmente essere al centro della carta di Smith; questo, in altre parole, è traducibile con l'aver il modulo molto vicino al centro della carta e che si può spostare di poco da esso, ma la fase enormemente variabile, dal momento che il modulo potrebbe spostarsi in ogni direzione, portando ad avere una fase enormemente variabile;
- il caso duale è per $|\Gamma| \sim 1$: il modulo in questo caso può cambiare enormemente (come si vedrà anche dall'espressione che verrà ricavata tra breve), mentre la fase subirà lievi variazioni.

Le curve, dunque, saranno non lineari rispetto al coefficiente di riflessione; curve di fase e curve di modulo sono una l'opposto dell'altro. Tutto ciò si vede sviluppando le equazioni rispetto a 0 e a 1 (e tra breve faremo anche lo sviluppo a 1). In pratica, ciò che si fa, è verificare la calibrazione mediante degli standard di verifica, *validation kit*.

Si fanno misure di altri oggetti; uno dei più utilizzati è l'offset short.

Proviamo a questo punto a sviluppare rispetto a 1 la famosa espressione; in questo caso non si può più utilizzare il risultato noto da McLaurin, dunque sarà necessario derivare l'espressione:

$$\left. \frac{\partial \Delta \Gamma}{\partial \Gamma_x} \right|_{\Gamma_x=1} = \frac{(1 - E_S \Gamma_x)(-E_S E_D + E_R - 1 + 2E_S \Gamma_x - (-E_S)(E_D - E_S E_D \Gamma_x + (E_R - 1)\Gamma_x + E_S \Gamma_x^2))}{(1 - E_S \Gamma_x)^2}$$

Dal momento che siamo interessati al primo ordine, tutti i termini quadratici o dati dal prodotto di due coefficienti (non coinvolgenti Γ_x) sono direttamente annullabili in quanto di ordine superiore. Possiamo dire che:

$$(1 - E_S \Gamma_x)[(E_R - 1) + 2\Gamma_x E_S] + E_S E_D + E_S(E_R - 1) + E_S^2 \Gamma_x^2 = \\ \sim \Gamma_x(1 + E_S)$$

Cosa significa ciò? Beh, il $\Delta \Gamma$ è circa uguale a 1, ma in più si ha il termine E_S , che rappresenta un vettore che gira attorno al punto interessato, sulla carta di Smith.

Questo vettore, ruotando, finisce per andare al di fuori della carta di Smith; esso è sostanzialmente derivante dalla presenza di un source match residuo.

Abbiamo la possibilità di collegare semplicemente source match residuo (il parametro nelle specifiche) e misure ottenute. Si consideri per esempio il seguente andamento:

Si vede che, in questo caso, si arriva ad avere 1 dB di ripple; tenendo presente che $\Gamma_x \sim 1$, si può dire che:

$$\Delta \Gamma = 10^{\frac{1}{20}} = 1,122$$

se a questo togliamo il 1 del Γ_x :

$$20 \log_{10}(\Delta \Gamma - \Gamma_x) = 20 \log_{10}(0,122) = -18,27 \text{ dB}$$

ossia, con -18 dB di source match, si ha un ripple di 1 dB. Supponendo di avere 0,1 dB di ripple, si vede che il source match migliora moltissimo:

$$\Delta \Gamma = 10^{\frac{0,1}{20}} = 1,01158$$

se a questo togliamo il 1 del Γ_x :

$$20 \log_{10}(\Delta \Gamma - \Gamma_x) = 20 \log_{10}(0,01158) = -38,73 \text{ dB}$$

si hanno dunque circa 40 dB di source match.

Si fa di solito un errore, quasi classico: se si fa la taratura, e si misura lo stesso thru utilizzato per tarare, l'unico errore deriva dalla ripetibilità e

non dall'accuratezza; si arriva a vedere dei source match di 80 dB; la verità è che una cosa del genere non ha assolutamente senso: quando si fa la taratura, dopo bisogna “cambiare la tara”, e “misurare un pesetto uguale al precedente, ma non usare lo stesso pesetto di prima”. Fare inoltre misure al di sotto del GHz è sostanzialmente impossibile, a causa della presenza delle enormi perdite che vanno a incrementare la parte immaginaria dell'impedenza portando a effetti sgradevolissimi. Il check va dunque fatto, ma tenendo assolutamente conto di questo accorgimento, al fine di avere un riscontro sensato.

Spesso si evita di fare misure di verifica di tipo “assoluto” (soprattutto a multiporte): dal momento che sarebbero veramente troppo pretenziose, si richiede sostanzialmente di avere due tipi di misure verificate.

- Misure di simmetria: si richiede che $S_{11} \sim S_{22}$; si parla di simmetria dal momento che, se questi parametri sono uguali, il 2-porte rappresentato dalla matrice scattering è simmetrico. Il fatto che queste misure siano quasi coincidenti è già buono: si tratta chiaramente di una condizione necessaria ma non sufficiente all'aver una buona misura, tuttavia è una buona base da cui partire per verificare la correttezza della taratura. È dunque fondamentale che alle due porte, al fine di garantire questa cosa, si utilizzino gli stessi standard: garantire la visione degli stessi carichi alle due porte, in modo da cercare di rendere questi parametri molto vicini. Fare ciò in certe situazioni, come quella del coassiale, può essere problematico: dovremmo avere lo stesso sesso per i connettori, dunque non si può fare (non sono APC7).
- Misure di reciprocità: si tratta di misure molto più importanti; quando si fa il plot di S_{12} e S_{21} , si pretende assolutamente che siano uguali, altrimenti la taratura sarebbe completamente sballata. I coefficienti di reciprocità dipendono solo dal parametro α , dunque il fatto che essi siano sbagliati porterebbe a capire che la taratura è completamente insensata. Se la condizione precedente era necessaria, questa lo è ancora di più.

Nel caso avessimo la possibilità di progettare un nostro personale set di standard, un suggerimento è quello di mettere due offset short (in quanto a banda stretta), una linea a 50Ω , una linea a impedenza caratteristica diversa (magari 100Ω per esempio), in modo da poter fare anche la validazione della calibratura.

3.6.5 Tecniche avanzate di taratura di multiporte

Nel caso dei network analyzer multiporte oltre a tutto ciò che è stato finora detto, si hanno altre caratteristiche aggiuntive: è necessario fare più connessioni, dunque potrebbe essere necessario fare più prove di ripetibilità, e verificare meglio l'accuratezza, dal momento che si potrebbero avere ulteriori peggioramenti sotto questi punti di vista. Come noto, gli standard a multiporte sono sostanzialmente standard a 2 porte, e tipicamente ciò che si cerca di fare per calibrare è usare la TRL.

Quando si parla di multiporte, però, spesso si parla di applicazioni molto avanzate, dunque ciò che si fa, al giorno d'oggi, è ricorrere a kit di calibrazione più complicati e, se possibile, personalizzati (se il misurista in questione sa quello che fa). Una configurazione per esempio utilizzabile potrebbe essere la seguente:

Il problema più grosso nei multiporte è il kit di verifica: i *verification elements*; nel caso del set presentato, si hanno, oltre al normale kit TRL, un *beatty standard* costituito da linee di impedenze caratteristiche diverse (una più larga dell'altra), una struttura a linee accoppiate, e un unknown thru (il *ragno*), ossia una struttura che collega tutto con tutto. Si può fare una calibrazione a 12 porte in questa maniera:

L'alternativa è la seconda: tante TRL, e poi mettere un unknown thru per metterle tutte assieme. Le restanti strutture possono essere poi usate per verifica.

Si può fare quello che si vuole, a patto però che le strutture siano tutte consistenti tra di loro: a gruppi di 4 e fra di loro.

Le verifiche che si fanno soprattutto sono misure di **consistenza**: $S_{11} = S_{22}$, verificare che il coupling abbia certi valori, e così via; queste misure sono, come già detto, meno stringenti di quelle effettive sui valori dei coefficienti da misurare, ma chiedere che queste misure vengano è già molto meglio che non chiedere niente: pretendere immediatamente che i valori vengano corretti è impossibile, mentre chiedere che si abbia una consistenza di questo tipo è già più ragionevole. Il consiglio è dunque quello di fare almeno le misure di consistenza. Con questi standard torna inoltre il problema dei piani di riferimento: non si ha idea di dove esso sia, poichè ci son alcune linee curve, alcune dritte, dunque molti elementi di indeterminazione.

Taratura su wafer

Quello della taratura su wafer (multiporte) è ancora un altro problema; per questo vale tutto ciò che è stato detto prima, più un altro: minimizzare i

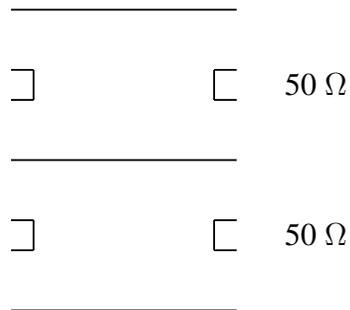
touchdown. Le probe non devono mai essere mosse, mentre il ciak, il porta-wafer, è l'elemento da muovere.

Generalmente, per questi motivi, la TRL non si fa: si dovrebbe mettere e togliere roba.

Il problema è l'averne a che fare con linee una vicino all'altra: è come avere quattro LRM, dove però, essendo le linee vicine, si rischia che esse parlino.

Perché la dynamic calibration funzioni, serve un path che colleghi tutte le porte; nella figura precedente, per esempio per andare dalla 5 alla 4, esiste questo path? Beh, osservando i vari standard a disposizione, se si riesce a trovare un percorso in grado di collegarli, allora siamo tranquilli. In questo esempio, dalla 5 vado alla 1, dalla 2 alla 2, dalla 2 alla 6, dalla 6 alla 7, dalla 7 alla 3, dalla 3 alla 4. Ciò vale ovviamente per qualsiasi porta. Questa cosa garantisce di avere, insieme ad altre condizioni, di avere un set di equazioni indipendenti.

Una soluzione ideale è quella di fare uno standard di questo tipo:



Questo può essere ottimo perché non fa parlare tra di loro le linee. Facendo shiftare questo standard, e aggiungendo una linea, si riesce a risolvere il problema di crosstalk e comunque a ottenere più path (5 invece che 4), realizzando due condizioni di connessione del LRM.

Sostanzialmente, il suggerimento è: progettare un circuito tale per cui si possano avere tutti gli standard necessari per calibrazione e verifica, e portare al punto desiderato i piani di riferimento.

Taratura su socket board

Si consideri un altro esempio di struttura piuttosto complicata da tarare: una socket board.

Questa è una cosa ancora più sofisticata. Quando si ha a che fare con qualcosa del genere, si hanno moltissime linee di accesso (i vari contatti): esse sono vicinissime, dunque esse **parlano** fra di loro. Per eliminare questo effetto è necessario fare un modello elettromagnetico della cosa, e de-embeddare il multiporte dal piano di riferimento.

Per misurare una struttura complicata come questa, è necessario definire un certo numero di passi.

1. Definire la struttura effettiva da misurare: ciò che ci permette di definire una sensitivity rispetto a ciò che stiamo cercando; data una struttura a 50 o 100 porte, e un VNA a 12 porte, bisogna capire quali sono le porte utili; fatto ciò, è necessario de-embeddare le linee di accesso; tutte le volte che si può, si metta il piano di riferimento della taratura direttamente all'interno delle linee di accesso, progettando standard con davanti la stessa linea di accesso.
2. Definire una procedura di taratura (se il software permette di fare il gioco della procedura migliore).
3. Costruire gli standard.
4. Verificare la calibrazione: fondamentale.
5. Fare la misura.

Due soluzioni (almeno):

Una prima soluzione (sinistra) è quella di utilizzare dei thru alle varie porte: una LRM, e diversi thru; si prende una configurazione di standard di questo tipo:

In questo caso, però, si ha il problema di prima: la presenza di crosstalk. Inoltre, sono state spostate le probe. Ciò porta a usare anche la seconda soluzione: LRM, e vari reciprocal per creare la fine del path, in modo da eliminare il crosstalk:

Ciò elimina il crosstalk, permette di non spostare le probe, e solo cambiando l'ordine degli standard: la potenza della dynamic calibration sta proprio nel fatto di poter usare gli stessi standard, in ordini diversi, e ottenere risultati diversi (in questo caso, tendenzialmente migliori: con l'accuratezza propria della LRM si ottiene l'eliminazione del crosstalk e si mantengono fisse le probe).

Nel caso il procedimento sia fatto bene, si possono ottenere buone misure di verifica della simmetria anche sui parametri differenziali, S_d : essi sono necessari, per la caratterizzazione dell'elettronica, come già detto, dal momento che si tende a trasmettere l'informazione all'interno dei modi differenziali (per evitare le perdite dei dielettrici), ma, essendo i parametri differenziali definiti per l'appunto mediante differenze, è facile che gli errori diventino molto elevati. Se la taratura è stata fatta bene, si riescono a ottenere similitudini anche a livello di -40 dB.

Si vuole a questo punto parlare ancora un poco di crosstalk: il far-end crosstalk.



Il far-end cross-talk è una misura di reciprocità: va a colpire porte tra esse lontane, che parlano mediante crosstalk (lontane per la progettazione della socket board); la misura deve portare a vedere che esse siano uguali. Le due porte, essendo lontane, comunque, si devono parlare poco.

Discorso diverso per il near-end cross-talk: i cross-talk tra porte vicine. Questi possono essere diversi, dal momento che derivano da porte vicine, ma su board diverse, dunque da board su substrati diversi, che possono risentire in modo diverso degli effetti della diafonia.

3.6.6 E-CAL: Electronic Calibration

Quando si parla di analizzatori di rete a due porte, esiste ancora una sorta di tecnica di calibrazione, o meglio di metodo per calibrare: il E-CAL, ossia il Electronic Calibrator: si tratta di moduli con dentro degli switch pin-diodes, e un certo set di standard, resistori, linee, che vengono precaratterizzati in fabbrica attraverso misure TRL. Si mette a questo punto in un file questa caratterizzazione, e quando si attacca al VNA questo blocco, dopo aver avviato la procedura automatica, il calibratore elettronico cambia gli switch e con essi gli standard, facendo poi i conti dei coefficienti. La taratura quasi sempre è una SOLT, e spesso non è molto buona; dal momento che si parla di pin diodes, la ripetibilità non è un grosso problema, ma la loro stabilità lo è. Come detto, questa strategia funziona per dei 2-porte, ma non si usano ancora per strumenti più complicati, a più porte.

Essendo un transfer standard, la sua accuratezza dipende dall'accuratezza della TRL con cui è stato calibrato, dall'accuratezza del VNA usato per calibrarlo, e cos via.

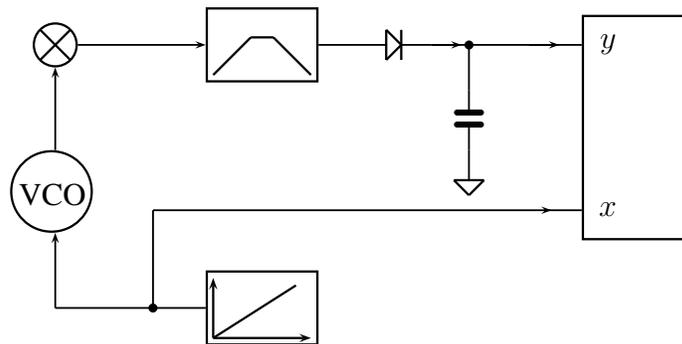
Capitolo 4

Introduzione ad alcuni strumenti di misura

4.1 Analizzatore di spettro

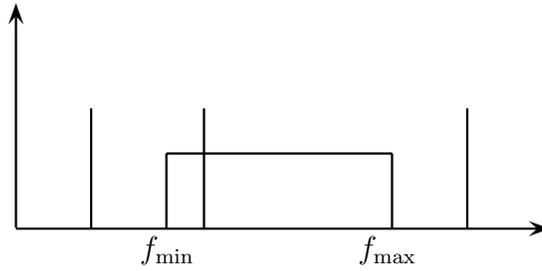
Questo capitolo della trattazione sarà dedicato allo studio di uno strumento diverso rispetto ai precedenti: l'analizzatore di spettro. Questo può essere realizzato in diverse maniere: a eterodina, con banchi di filtri, mediante FFT. A microonde, l'unico metodo che di fatto si può utilizzare è l'eterodina: esistono sampling scope a decine di GHz però la FFT potrebbe portare comunque a troppe complicazioni.

Detto che dunque lo schema di base sarà quello fondato sull'eterodina, possiamo abbozzarlo:



Si ha una rampa di tensione che andrà a comandare da un lato il VCO, dall'altro l'asse x visualizzato; si avrà quindi un mixer che andrà a far battere ingresso e VCO, e quando il battimento avrà la stessa frequenza del filtro passa banda un rilevatore di picco potrà andar pilotare l'asse y dell'uscita.

A questo punto è necessario determinare una cosa: dove possiamo mettere la frequenza intermedia, IF? Beh, vediamo un po' il seguente disegno:



Di sicuro non potremo mettere la IF all'interno della banda del segnale: se $f_{IF} = 5$ GHz, e la frequenza di ingresso $f_{in} = 5$ GHz, la frequenza del VCO dovrebbe essere a 0 Hz, cosa insensata.

Seconda idea potrebbe essere quella di mettere la frequenza intermedia prima di f_{min} , ossia della frequenza più bassa della banda; in questo caso, si avrebbe:

$$f_{IF} = f_{in} - f_{VCO}$$

ma, d'altra parte, i battimenti sono due, dunque si potrebbe anche avere:

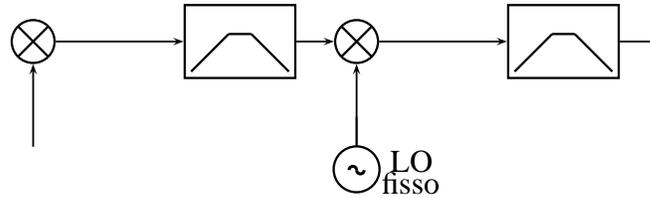
$$f_{IF} = f_{VCO} - f_{in}$$

in questo modo, nel caso per esempio in cui $f_{IF} = \frac{f_{in}}{2}$, si avrebbe che la frequenza del VCO, f_{VCO} , dovrebbe essere:

$$f_{VCO} = f_{in} + f_{IF} = \frac{3}{2}f_{in,min}$$

Questo significa che il VCO potrebbe andare a mixarsi con i prodotti sia superiori sia inferiori, ottenendo dunque delle immagini, che non possono essere cancellate.

Andando per esclusione, il punto in cui è necessario mettere la frequenza IF è a destra, ossia per frequenze più alte della frequenza massima; non vi saranno frequenze immagini, a patto che si introduca un filtro passa basso (a frequenza sufficientemente elevata, nella fattispecie maggiore o uguale alla f_{max}) prima del mixer. Questo procedimento purtroppo però non è indolore: noi dobbiamo mettere il VCO a frequenze superiori alla f_{max} ma, quando siamo a microonde, questa potrebbe essere anche a frequenze altissime; questo introduce la necessità di avere più di uno stadio di conversione, in modo da avere più di uno step. Volendo dunque andare da pochi Hz a decine di GHz è necessario usare un'architettura a doppia o tripla conversione. Si ha uno schema di questo genere:



La terza conversione, ove presente, spesso è numerica, dunque attuata sul dato già digitalizzato (dopo il ADC). Spesso la conversione potrebbe anche essere fatta su più bande (la prima): si usano mixer e filtri diversi per bande diverse, e anche VCO; i pin diodes sono invece dispositivi generalmente a banda larga, dunque non hanno problemi.

A questo punto, una domanda: da cosa dipende la velocità dell'analizzatore di spettro? Beh, come detto, tra i vari blocchi presenti nell'architettura vi sono anche dei filtri; questi filtri devono essere selettivi, dunque il ringing potrebbe durare molto tempo, e i filtri potrebbero anche introdurre un delay. Il fatto che il filtro sia molto selettivo infatti coincide con la presenza di molti poli, di molti elementi reattivi; se vi sono molti poli, dunque, vi saranno molti elementi di rallentamento e di introduzioni di oscillazioni meno smorzate: velocità e risoluzione sono in competizione.

4.1.1 Analisi del FSU Rohde & Schwarz

SI vuole a questo punto analizzare, a partire dal manualetto annesso, uno strumento reale: il FSU di Rohde & Schwarz. Verrà utilizzato un elenco numerato, dove il numero riferisce alla pagina (in ordine).

1. SI faccia riferimento al diagramma a blocchi: si può vedere che si ha un blocco composto da filtro YIG, Diplexer e convertitore a microonde; lo YIG filter serve per fare una preselezione, ossia per pulire il segnale di ingresso in modo che al mixer arrivi solo lo spettro utile, evitando le immagini; questo YIG rappresenta dunque il ruolo del filtro passa basso di preselezione, da porre prima del mixer, con l'unica variante rispetto allo schema concettuale dell'essere YIG, dunque tunabile. Il sistema è a tripla eterodina.
2. Il funzionamento del sistema è il seguente: prima di tutto si ha un attenuatore a step di 5 dB da 0 a 75 dB, che può ridurre il livello del segnale prima di farlo entrare nei moduli a RF; nel caso si abbia a che fare con un modello ad alte frequenze (maggiori di 3,6 GHz) si fa passare la RF per un diplexer prima di farlo entrare nel convertitore a microonde, altrimenti va direttamente nel convertitore. Il modulo IF è

lo stesso in tutti i modelli. L'attenuatore è a scatti, a celle, fino a 75 dB, per fare in modo che la dinamica di ingresso del segnale (il range di ampiezze inseribili) sia il più ampio possibile; il mixer, per poter sentire il segnale, dovrà essere composto da diodi piccoli, che però essendo piccoli potranno portare correnti ridotte: i segnali che dovranno entrare nei diodi dunque non dovranno essere troppo alti o troppo bassi. Alcune altre note: si ha, in ingresso all'attenuatore, anche un segnale a 128 MHz (il CAL-INPUT), rappresentante il riferimento di frequenza. Ultima nota riguardo il condensatore: i modelli a frequenze relativamente alte (8 e 26 GHz) hanno un condensatore di accoppiamento in AC, che può essere bypassato; i modelli da più di 26 GHz sono permanentemente accoppiati in DC, dal momento che il condensatore deve essere introdotto dall'esterno (bisogna entrare con un condensatore): il modello di un condensatore infatti come noto è composto da resistenze, capacità, induttanze, dunque il condensatore non lavorerebbe da capacità a quelle frequenze, e si un accoppiamento permanente in DC.

3. Blocco di conversione per sistemi che lavorano con meno di 3,6 GHz: si ha qualcosa di simile al principio di prima, ossia si ha una RF, un passa-basso, un mixer doppio bilanciato. Come già detto, si han due conversioni, per riportare il segnale da 20 Hz a 3,6 GHz in ingresso, a una IF di 20,4 MHz. Ciò accade in due step:
 - il primo mixer è realizzato con un oscillatore a YIG, da 4628 a 8229 MHz; fuori dal mixer si han amplificatori a bipolari (meno suscettibili al rumore flicker); si ha un segnale in uscita bilanciato, in modo da evitare di utilizzare un trasformatore a presa centrale in più, il quale introdurrebbe perdite nel processo di conversione di frequenza. Si ha dunque un secondo mixer al quale si entra direttamente col segnale bilanciato, dal momento che convertire il segnale in single-ended e ri-convertirlo in bilanciato non ha senso (e si passa per un amplificatore da 11 dB, che sarà a sua volta bilanciato); la cosa inoltre riduce i prodotti di intermodulazione del mixer;
 - il secondo mixer farà la conversione di frequenza per il segnale che dovrà entrare alla seconda conversione vera e propria. In uscita da tutto si ha un trasformatore a frequenza fissa, a 400 MHz, che rende il segnale bilanciato single-ended per la conversione alla seconda IF;
 - nella seconda IF si incomincia a fare il processing del segnale in termini di definizione della resolution bandwidth (RBW): a sec-

onda di quanto il filtro è stretto si ha una risoluzione spettrale più o meno fine; si fa filtraggio mediante due filtri a cinque stadi di risonatori ceramici, ossia filtri basati su elementi realizzati su substrati ceramici (come quarzo): si tratta di filtri a elevatissimo Q .

- Quando si ha a che fare con uno span di 30 MHz, i prodotti di intermodulazione di secondo ordine del terzo mixer potrebbero diventare importanti, e creare dei problemi; ciò che si fa è usare l'attenuatore in modo da ridurre la potenza del segnale in ingresso al mixer, se il livello di ingresso è maggiore di - 25 dB_m. Si ha dunque, dopo il terzo mixer (quello posto dopo il suddetto attenuatore) il segnale IF3 OUT, ossia la terza conversione di frequenza.
4. Se si ha a che fare con sistemi ad alta frequenza, ossia che lavorano con più di 3,6 GHz, il primo pezzo con cui bisogna avere a che fare è lo YIG filter: qua infatti si deve avere anche una parte atta a convertire la microonda. Si ha qualcosa del genere:
- Una parte del segnale (dalla DC ai 3,6 GHz) viene mandata direttamente al convertitore RF prima analizzato;
 - Per le frequenze più elevate, ovviamente non è possibile avere frequenze generate dal VCO troppo elevate, dunque quello che si fa è sfruttare i prodotti di intermodulazione, usando $f_1 - Nf_{VCO}$, dove N dipende dal modello o dalle frequenze da gestire. Lo YIG filter permette di fare preselezione di una porzione di banda molto ben definita, e dunque di scartare la porzione di banda che non ci interessa, dunque evitare di avere immagine nonostante si ponga la frequenza del VCO prima di quella minima. Il mixer è un mixer armonico, nel senso che sarà predisposto per lavorare con i termini a frequenze più alte.
 - Per i modelli sopra i 26 GHz si fa una conversione doppia: si ha una opzione che fa fare una seconda conversione per fare abbassare ulteriormente la frequenza: si esce sempre con una frequenza di uscita di 4 GHz, e si cerca di arrivare al 404 finale, per poter arrivare sempre al convertitore RF.
5. Il filtraggio a IF è a bassa frequenza (rispetto alle microonde: si tratta di un filtraggio dopo la terza IF). Al giorno d'oggi questa parte si potrebbe probabilmente fare in digitale, dopo aver usato un campionario a 100 MHz e lavorando con software radio. Ciò nonostante, è buona cosa

sapere come funziona il processing, dal momento che conoscere come fare i filtri è sempre importante anche in vista di un passaggio al software radio. Si ha la possibilità, in questo strumento, di avere bande per la risoluzione da 1 Hz a 20 MHz, in step 1/2/3/5. In ingresso a tutto si ha un multiplexer che divide le varie bande: per bande da 100 kHz a 50 MHz si usano cinque banchi di filtri LC; per bande superiori a 5 MHz si passa dal filtro wide band; per bande tra i 10 Hz e i 100 kHz si utilizza una conversione in digitale, con una condizione in più: il fatto che, per frequenze inferiori a 30 kHz, si usa il filtro a cristalli; questo è un filtro tunabile da 2,5 kHz a 70 kHz, e la sua funzione è quella di non far entrare nel ADC troppo rumore o troppa informazione (avere 20 MHz di informazione all'interno del ADC infatti lo porterebbe in saturazione).

6. Il processing digitale, come detto in precedenza, riguarda le frequenze inferiori ai 100 kHz: per esse la RBW viene generata mediante un DDC (Digital Down Converter). La banda gestita è complessa: questo spectrum analyzer in realtà è un prodotto particolarmente avanzato dal momento che non rileva solo il modulo come i suoi predecessori classici, bensì fa anche una rilevazione di fase, diventando di fatto un Vector Signal Analyzer. Si usa un NCO come altro termine di mixing per il segnale di ingresso del DDC, dunque il segnale I/Q viene mandato a un filtro HDF (High Decimation Filter) e a un FIR.
7. L'analizzatore di spettro ha ancora le bande video (VBW) e detectors. Come si era visto nello schema di base, si ha un rilevatore di picco, detto "filtro video", dal momento che esso media il segnale, prima di presentarlo sullo schermo.

La differenza rispetto agli analizzatori vecchi è la possibilità di avere dei peak detectors, ossia dei voltmetri di picco, a valor medio, RMS. Quasi tutti gli analizzatori di spettro moderni hanno di queste funzionalità, che permettono di emulare i suddetti voltmetri.

Al giorno d'oggi gli analizzatori di spettro sono dunque macchine molto più potenti, che permettono di avere molte più funzionalità (come scritto nell'ultimo punto) rispetto alla banale visualizzazione del modulo della trasformata di Fourier.

4.2 Power meters

4.2.1 Introduzione

In questo capitolo si vuole introdurre una serie di concetti atti a misurare, mediante strumenti diversi dall'analizzatore di spettro (la cui accuracy è assolutamente poco valida, trattandosi di strumenti principalmente atti a effettuare misure qualitative), la potenza a frequenze elevate.

Quando si parla di misurare potenze, si può sostanzialmente parlare di due tipi di queste grandezze: average power, ossia potenza media, o peak power, ossia potenza di picco (massima potenza in un certo intervallo di tempo); a seconda che si voglia misurare l'una o l'altra potenza, si dovrà avere a che fare con sensori adatti a una misura o all'altra.

La potenza di picco è molto importante dal momento che, quando si ha un certo sistema, è molto probabile che alcuni blocchi, per una ragione o un'altra, abbiano degli overshoot, ossia delle sorte di "sovralongazioni" della potenza; può essere fondamentale conoscere l'entità di questi, al fine di sapere come il sistema dovrà lavorare al fine di gestire questo genere di problemi. Altro esempio in cui trattare la peak power è fondamentale è quello dei radar: si tratta di applicazioni in cui si ha a che fare con dei kW di potenza impulsati, dunque con potenze elevatissime.

In trasmissione parlare di potenza di picco è generalmente meno importante: si parla di riferibilità, (come vedremo tra breve) ossia di poter riportare una misura con valori noti a campioni appropriati, mediante confronti, ma il transistor funziona in commutazione, in switching, dunque spesso quella che conta è la potenza media. La potenza media è molto importante anche dal momento che essa è l'unica che si riesce a riferire bene fino alle microonde. Definiamo dunque la potenza media come:

$$P_{\text{avg}} = \frac{1}{nT} \int_0^{nT} e(t)i(t)dt$$

Dal momento che i power meter a microonde hanno costanti di tempo molto più lunghe di quelle delle onde trattate, essendo le frequenze delle microonde molto elevate, a microonde si fa "davvero" la media.

Parlando di commutazione, di switching, spesso si ha a che fare con impulsi a onda quadra; ha senso chiedersi quale sia la potenza associata all'impulso, dato il suo duty cycle D : la pulse power P_p è definibile come:

$$P_p = \frac{P_{\text{avg}}}{D}$$

Tutte queste definizioni vanno bene, ma solo a regime: esse non considerano in alcun modo la presenza di eventuali overshoot, dunque dei transitori: si fa la potenza media, la si divide per il duty cycle, e tutto è finito.

Una definizione più sofisticata potrebbe essere quella della peak envelope power: quando le forme di modulazione sono diventate molto più complesse, è stato necessario parlare di questo tipo di definizione, al fine di tenere conto del contenuto spettrale dell'oggetto.

Volendo fare la potenza media dovremmo prendere una porzione dello spettro e farne l'integrale; quello che si fa in questo tipo di applicazioni è o andare a vedere il rapporto spettrale tra una certa parte del segnale e la potenza media, come nel ACPR, oppure definire nel dominio del tempo un gate e misurare per il periodo di tempo ad esso associato la potenza.

4.2.2 Power meter a microcalorimetro

Come si realizza la riferibilità di questo genere di strumenti? Sostanzialmente, l'idea è quella di partire da uno standard primario: un microcalorimetro. Per microcalorimetro si intende un sistema termico molto accurato dove si valuta la potenza a microonde a partire da un confronto con la potenza in DC. Ciò che si fa coi power meter è vengono però confrontarli con una testina a sua volta tarata mediante un microcalorimetro. I power meter, dunque, sono sostanzialmente dei bolometri, applicati a un qualche ponte (il più classico è il ponte di Wheatstone), al fine di capire come vari lo sbilanciamento al centro del ponte. Tutti i power meter sono averaging power meter.

Nel microcalorimetro si calcola con estrema accuratezza la potenza in DC che viene trasferita al bolometro, e si fa la comparazione con la potenza a microonde che porta alla stessa variazione di tensione al ponte, e tutto ciò in un ambiente termico estremamente ben stabilizzato. Una volta invece esistevano i calorimetri, ossia un pentolone in cui si mandava della microonda, la quale scaldava l'acqua, e con la temperatura dell'acqua misurata si riesce a determinare quanta potenza sia stata mandata.

Al fine di sentire la differenza si utilizza una termocoppia: si tratta di un'alternativa alla misura mediante ponte di Wheatstone: il bolometro è collegato a una termocoppia la quale scaldandosi fa variare la tensione a vuoto misurata.

Uno schema interessante in grado di modellare il bolometro è il seguente:

Le due resistenze sono messe nello stesso blocco termico, in modo da poter avere la stessa escursione termica; poi, solo una delle due subirà la microonda, che modificherà il valore della resistenza.

Lo schema più interessante probabilmente è questo:

Da un lato si ha il ponte a RF (RF bridge), il quale è il normale, solito ponte di Wheatstone; si ha poi un operazionale che effettua il sensing della differenza al centro del ponte, e, dunque, esso fa in modo da imporre a 0 la differenza di temperatura al centro del ponte (essendo esso in linearità). L'altro ponte è il ponte di compensazione, che produce una tensione di compensazione v_c ; dall'altra parte (dal ponte a RF) si ha in uscita la tensione v_{rf} . Gli operazionali sono tali da auto-bilanciare il ponte: questo permette di avere la compensazione, grazie al termistore (l'elemento critico). Il fatto di avere queste due parti permette di ridurre il modo comune.

La parte "chopping and summing" è quella che permette, data la tensione a RF, e la differenza con quella dipendente dalla sola temperatura, di fare la somma e la differenza delle due; questo verrà poi chopato (come nella slotline, per rendere il segnale più semplice da condizionare, da amplificare). Si fa la differenza dei due segnali chopati, la si amplifica, dunque si ri-choppa il segnale con un impulso proporzionale alla somma.

La formula del bolometro è:

$$P_{rf} = \frac{V_{rf0}^2}{4R} - \frac{V_{rf}^2}{4R}$$

In questo modo si vede qual è il cambio rispetto al valore iniziale a RF, V_{rf0} . Ciò si può scrivere come:

$$P_{rf} = \frac{1}{4R} (V_c - V_{rf}) (V_c + V_{rf})$$

Si moltiplicano due segnali, uno impulsato proporzionale alla somma delle tensioni, l'altro alla differenza, e così si ottiene l'informazione desiderata.

4.2.3 Power meter a termocoppia

I power meter basati sulle termocoppie sono quelli più utilizzati: si realizzano termocoppie integrate, come la seguente:

Si deposita sul biossido di silicio del materiale come il nitrato di tantalio per fare la giunzione, e dunque si costruisce un qualcosa che scalda una resistenza integrata e che, scaldandosi, varia la tensione. Si ha in ingresso un coassiale, con un adattatore tapered, e si esce con dei μV per grado celsius (di solito, 250).

Si è passati dall'usare termistori (con ponti) all'usare massicciamente la termocoppia: questo sostanzialmente dal momento che si ha un transitorio molto più breve, quando si usa la termocoppia.

Lo schema di base di un power meter a termocoppia è il seguente:

Si han termocoppia, chopper (essendo il segnale piccolo), e il meter. Si noti che i due blocchi “sensor” e “meter” sono distinti: questo è interessante dal momento che, al giorno d’oggi, molto spesso sensori e power meter sono diversi, in scatole diverse, e ciò è ottimo. I power meter inoltre non sono più particolarmente utili, dal momento che molti sensori escono direttamente in USB, e mediante un ADC e un po’ di condizionamento è possibile collegare direttamente il sensore al computer.

Quando si parla di accuratezza, il colpevole di sicuro sarà il sensore: che sia USB o che sia scatolotto power meter esso non introduce niente di chè. Le testine sono piccole, dal momento che, per avere minor inerzia termica e meno parametri parassiti, è meglio avere a che fare con resistenze piccole; queste, di conseguenza, potranno gestire poca potenza (al massimo sui 100 mW). Ciò che si fa dunque per aumentare la potenza che si può introdurre nel sensore è anteporre degli attenuatori: ciò si vede quando i sensori sono molto lunghi: si han degli attenuatori che allungano fisicamente il dispositivo.

4.2.4 Power meter a diodi

Il terzo tipo di sensore sono, banalmente, i diodi: i diodi sono elementi non lineari ma, di fatto, viene usato solo dal ’74 per misurare nel senso vero e proprio della parola delle potenze, dal momento che un diodo è quadratico ed è veloce. Si ha uno schema di questo genere:

Questo è un esempio di power meter a diodi, e la cosa bella, oltre la velocità, è il dynamic range: si parla di 90 dB.

In sostanza, si ha un ponte, choppato, amplificato, e in un chip si introducono i due diodi, uno per la semionda negativa e uno per la positiva, collegato al resto del circuito. Il diodo è un ottimo sensore di temperatura: vedendo la caratteristica del diodo si ha la dipendenza da $k_B T$, e mediante un fitting logaritmico dei dati si riescono ad avere curve molto valide.

Wide-dynamic-range average power sensors

Una cosa che sempre più spesso diventa importante, con la nascita delle nuove modulazioni, è quella di effettuare misure accurate in presenza di fattori di cresta elevati (ossia presenza di una PAPR, peak-to average ratio elevata). Servono sensori con più di 50 dB di dynamic range, e quindi power meter come questi non possono soddisfare ciò. Se il segnale a microonde cambia di 50 dB, si può misurare la potenza media ma non si può apprezzare sicuramente la presenza di picchi di potenza, e ciò con i power meter per ora visti non si può fare. Ciò che si fa allora è qualcosa di più avanzato, come ciò:

In questo modo si aumenta la linearità del sensore.

4.2.5 Accuracy dei power meter

L'accuratezza può derivare da diversi fattori, da diversi eventi. Un evento potrebbe essere quello di avere disadattamento, ma esso è un errore sistematico, e in quanto tale è possibile rimuoverlo per l'appunto adattando. Il disadattamento è un esempio ma assolutamente valido: esso è la massima causa di errore sistematico nei dispositivi.

Per ridurre i vari errori sistematici, è necessario effettuare un processo di taratura. Questo è sostanzialmente basato sui seguenti elementi:

- zeroing (azzeramento): si tratta di un “azzeramento” del sistema, in modo da far coincidere il livello di riferimento con uno zero; il livello di riferimento potrebbe per esempio essere quello tale per cui non si ha segnale a RF in ingresso.
- CAL-factor: si tratta di un coefficiente di taratura, la cui principale causa è il disadattamento. Questo coefficiente è dipendente dal termine $[1 - |\Gamma|^2]$, ossia dalla potenza riflessa. Generalmente, al giorno d'oggi, questo coefficiente è direttamente salvato nella testina.

Ciò che al giorno d'oggi dà ancora problemi è sostanzialmente il sensor. Si osservi, tornando al discorso del mismatch, il seguente grafico:

Come si può vedere, avere 0,1 dB di errore sulla potenza misurata è abbastanza normale: questo ci porti a capire che, dei bit di risoluzione, solo pochi saranno significativi, e che dunque prendere cifre al di sotto della prima, dopo la virgola, è tendenzialmente insensato.

Un parametro molto importante è l'efficienza del power meter, η_e , definibile come il rapporto della potenza netta assorbita dal sensore durante la misura e la potenza di sostituzione, a bassa frequenza, ossia la potenza che a bassa frequenza dà lo stesso effetto della potenza a radiofrequenza.

$$\eta_e = \frac{P_{\text{sub}}}{P_{\text{rf}}}$$

Di fatto, si ha una sostituzione, ma non è detto che la potenza assorbita e quella sostituita diano lo stesso effetto. L'efficienza del sensore rappresenta la differenza tra l'effetto della potenza di sostituzione e quella effettivamente assorbita dal sensore. Questo numero fa parte dei contributi che andranno a influenzare il CAL-factor, ed è da determinare a livello di standard primario, calcolando accuratamente le perdite.

Questa efficienza η_e è sostanzialmente influenzata da una incorretta distribuzione del calore a radiofrequenza piuttosto che in DC: l'elemento sensibile per quanto sia integrato, ha delle dimensioni fisiche; in DC questo

elemento è sostanzialmente puntiforme, rendendo la distribuzione del calore costante, uniforme. A RF è come avere tante resistenzine, ma dunque si rischia di non aver corrente uniforme su ogni punto, e dunque si potrebbe avere un riscaldamento non uniforme delle suddette; è possibile che vi siano degli “hot spot”, ossia dei punti in cui si abbia una variazione di temperatura diversa, maggiore, rispetto agli altri punti. Questo è un primo effetto, ma non l’unico: il Γ del power meter infatti si basa sull’assunzione che nulla cambi, ma non è vero: nei casi sofisticati si evita ciò, ma non è detto.

Le misure di potenza sono le misure **assolute** più accurate che si riesca a fare, e comunque sono poco accurate. Si può vedere che, essendo l’elemento non lineare, varia la resistenza, ma dunque con essa varia anche il coefficiente di riflessione, ottenendo problemi di disadattamento. Si hanno poi errori di zero set. Il numero più importante è l’incertezza in dB: 0,4 dB. Questo significa che non si riesce a misurare correttamente più di 0,4 dB. L’accuratezza migliore del NIST è di 0,15 o 0,12 dB, ma questo è il riferimento: al massimo si riuscirebbe a fare ciò con i riferimenti controllati dei centri di misura, quindi noi possiamo essere contenti con questi 0,4 dB.

4.3 Noise Figure meter

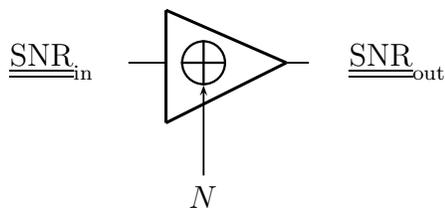
4.3.1 Richiami sul rumore

Prima di introdurre i metodi per la misura del rumore, vogliamo proporre una breve introduzione sul concetto di rumore. La quantità di solito utilizzata per determinare le prestazioni di un sistema sotto il punto di vista del rapporto segnale/rumore, ossia del rapporto della potenza di segnale sulla potenza di rumore. In altre parole:

$$\underline{\underline{\text{SNR}}} = \frac{P_S}{P_N}$$

Di solito, tutti i rumori presenti nel sistema sono considerati bianchi, ossia a densità spettrale di potenza costante.

Molto spesso capita di avere a che fare con dispositivi attivi, nella fattispecie per esempio con degli amplificatori. Cosa succede in un amplificatore? Beh, esso sostanzialmente aggiunge rumore:



È come se all'interno dell'amplificatore in qualche punto vi fosse un "nodo sommatore" nel quale si va ad aggiungere, al segnale, una componente di rumore. Si può dire che il rapporto segnale/rumore in uscita da un amplificatore sia un qualcosa del tipo:

$$\underline{\underline{\text{SNR}}}_{\text{out}} = \frac{G P_{\text{in}}}{G N_{\text{amplificatore}} + G N_{\text{in}}}$$

Ossia, il rapporto della potenza in ingresso, moltiplicata per un opportuno guadagno, e delle potenze di rumore che si hanno nel sistema: una sarà la potenza di rumore in ingresso all'amplificatore, moltiplicata per il guadagno, e una sarà un'altra potenza, che sarà quella "sommata nel nodo sommatore". Per semplicità di notazione, si utilizza una potenza normalizzata rispetto al guadagno dell'amplificatore G , in modo da poter avere un'espressione di questo tipo. Ovviamente, per quanto riguarda il rapporto segnale/rumore all'ingresso dell'amplificatore, esso sarà semplicemente quello di prima, dove però non si ha la componente aggiunta dall'amplificatore (e le varie componenti non sono moltiplicate per i guadagni, che tanto comunque vanno a semplificarsi come si può vedere nell'ultima equazione scritta). Si ha dunque:

$$\underline{\underline{\text{SNR}}}_{\text{in}} = \frac{P_{\text{in}}}{N_{\text{in}}}$$

A questo punto, può essere interessante calcolare la cifra di rumore come il rapporto tra il rapporto segnale/rumore in ingresso e quello in uscita:

$$\begin{aligned} F \triangleq \frac{\underline{\underline{\text{SNR}}}_{\text{in}}}{\underline{\underline{\text{SNR}}}_{\text{out}}} &= \frac{\frac{P_{\text{in}}}{N_{\text{in}}}}{\frac{P_{\text{in}}}{N_{\text{in}} + N_{\text{amplificatore}}}} = \frac{N_{\text{in}} + N_{\text{amplificatore}}}{N_{\text{in}}} = \\ &= 2 + \frac{N_{\text{amplificatore}}}{N_{\text{in}}} \end{aligned}$$

Nel caso poi il rumore sia bianco, è possibile scrivere ciò in termini di temperature equivalente di rumore (altro parametro, alternativo alla cifra di rumore, per la determinazione delle prestazioni sotto il punto di vista del rumore):

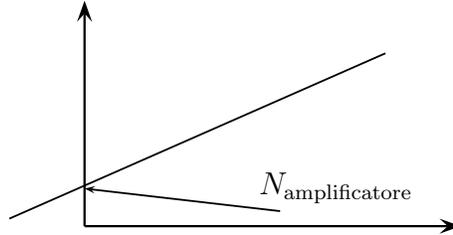
$$= 1 + \frac{T_{\text{eq,DUT}}}{T_{\text{eq,in}}}$$

Ricordando che $N = k_{\text{B}}TB$, è possibile usare T per determinare la temperatura equivalente di rumore.

Data questa equazione, possiamo fare il seguente ragionamento: se mandiamo in ingresso un rumore, possiamo dire che:

$$F - 1 = \frac{T_{eq}}{T_{eq,in}}$$

Quello che si può fare è scrivere l'espressione della potenza di rumore in uscita in funzione di quelle in ingresso. Si avrebbe un andamento di questo genere:



L'intercetta con l'asse è $N_{amplificatore}$. Volendo utilizzare la temperatura di rumore, si ha:

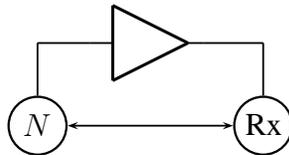
$$T_{eq,out} = T_{eq,in} + T_{eq,DUT}$$

ossia, posso dare in ingresso due temperature di rumore T_1 e T_2 (equivalenti dunque a due segnali rumorosi), fare l'interpolazione, misurare il rumore di uscita, quindi estrapolare il valore passante per 0, e questa sarà la temperatura di rumore aggiunta dal solo blocco attivo. Una volta nota la temperatura equivalente di rumore, è possibile rapportare questa a una temperatura di riferimento, quale per esempio la temperatura ambiente (che potrebbe essere per esempio pari a 25 °C, e dunque ottenere:

$$T_{eq,out} = 1 + \frac{T_{eq,DUT}}{25 \text{ °C}}$$

4.3.2 Hot-cold method

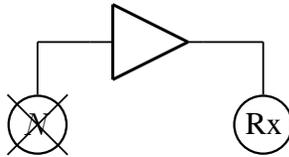
A questo punto, come si fanno le misure? Beh, l'idea è sostanzialmente quella di utilizzare una sorgente di rumore, la quale dunque produce rumore elettrico con spettro più o meno bianco, generante una temperatura di rumore equivalente a quella che avrebbe una resistenza riscaldata a 10000° C. Il generatore di rumore viene acceso e spento, e così via: hot e cold. In uscita dall'amplificatore si avrà un ricevitore per il rumore.



Il noise figure meter è un ricevitore che come unica funzione ha quella di controllare il generatore. Il generatore si può realizzare mediante un avalanche diode, ossia un diodo a valanga, caratterizzato dal parametro “ENR” (Excess Noise Ratio). Purtroppo, la cosa negativa di questo parametro è il fatto che esso varia in funzione della frequenza f che si considera. Per ENR si intende il rapporto tra il rumore quando il diodo è acceso e quando è spento.

Ancora due parole sulla caratterizzazione del rumore: nel caso di un transistor, la misura della cifra di rumore non è sufficiente al fine di effettuare una caratterizzazione completa; nel caso dei transistori, si usa il modello lineare, nella fattispecie i parametri scattering. Di solito si considera tutto adattato: come guadagno non si mette il S_{21} , ma il MAG. Nel caso dei transistori, oltre alla cifra di rumore serve avere il $\Gamma_{in,noise,opt}$: si tratta di quel coefficiente di riflessione da proporre al transistor in ingresso in modo da avere la cifra di rumore minima possibile. Si ha inoltre un terzo parametro, ossia la “resistenza equivalente di rumore”, parametro reale che deriva dagli altri al momento di determinare il modello.

Cosa possiamo dunque fare? Beh, sarebbe necessario introdurre un coefficiente Γ_{in} variabile sull’ingresso e, sulla carta di Smith, vedere come con l’impostazione del tuner cambia la cifra di rumore, utilizzando il metodo hot-cold. Si può però applicare un trucco, e togliere il generatore di rumore N , ossia il diodo:



Di fatto il Γ_{in} che si utilizza per realizzare la sorgente con il coefficiente di riflessione variabile introduce anche una certa porzione di rumore, dunque il tuner da solo potrebbe fare “da solo” da sorgente di rumore; questa variante viene detta “metodo cold”. Si noti che questa cosa si può fare se e solo se conosco il rumore aggiunto dal ricevitore stesso: anche il ricevitore infatti ha i propri problemi! Fondamentale diventa dunque, sotto questo punto di vista, la calibrazione del ricevitore. Tutto ciò che è stato studiato viene applicato a sistemi “banali”, nel senso di sistemi adattati: tutto a 50Ω .

Le misure che dobbiamo fare sono misure di potenze di rumore, dunque di sicuro niente di banale. Cosa ci salva? Sostanzialmente, il fatto che quello che ci serve è un rapporto tra potenze definito rispetto a una temperatura assoluta. Sarebbe possibile calcolare il rumore aggiunto di un dispositivo, se non vi fosse la parte attiva: dato un attenuatore, sappiamo esattamente dire

quanto esso aggiunge il rumore: sappiamo quant'è infatti la potenza che si dissipa! Con un transistoro questo ragionamento non è così facile da fare, ma la matrice scattering, o meglio la sua parte resistiva, è più o meno quella che fa rumore. Purtroppo misurare ciò che i transistori fanno sotto il punto di vista del rumore è praticamente impossibile: ciò che si può fare è misurarne i parametri scattering, dunque a partire da essi determinare una **stima** della cifra di rumore; si ribadisce che essa non deriva da una misura diretta, ma da una stima a partire da un'altra misura.

4.3.3 Strumento per la misura del rumore: il noise figure meter

Al giorno d'oggi la misura della cifra di rumore come detto si fa mediante stima, per esempio mediante un PNA-X, o con un analizzatore di spettro. Un tempo si utilizzavano strumenti specializzati in misure di questo genere.

All'ingresso del sistema si ha un range di frequenze accettato dai 10 ai 1500 MHz; se si hanno frequenze superiori, si mette un convertitore. L'ingresso di questo sistema è sostanzialmente uguale a quello di un analizzatore di spettro, con davanti un amplificatore. Controllare bene la cifra di rumore del primo amplificatore della catena (il "input amplifier") è fondamentale, dal momento che essa determinerà buona parte della cifra di rumore complessiva. Andando avanti nella catena si ha dunque un mixer, uno YIG filter, e la frequenza della IF è maggiore della frequenza di ingresso, proprio come in un analizzatore di spettro. Si ha dunque una seconda eterodina, a 300 MHz, una terza eterodina, a 20 MHz, e il noise power detector. La IF bandwidth finale, è di 5 MHz (ampiezza del filtro passa banda finale, ossia della B della formula $k_B T B$). Si ha quindi un voltmetro a singola integrazione, e dunque la misura della frequenza. Dalle specifiche si può vedere che si ha un range di 30 dB, resolution di 0,01 dB, incertezza di 0,1 dB.

Qual è la differenza tra questo strumento e un analizzatore di spettro? Sostanzialmente, il fatto che per questo non è possibile variare la banda IF, la quale è fissa ed è pari a 5 MHz. In questo strumento inoltre si ha il noise source drive output, al fine di avere la possibilità di attivare o meno il noise source, e quindi poter applicare il metodo hot/cold. L'input power detector serve ad avere un'idea della potenza, in maniera da regolare il controllo del guadagno ed evitare la saturazione del mixer. Il motivo per cui il filtro è a banda larga (questi 5 MHz) è per avere un po' di potenza, dunque "mediare meglio": più la banda è larga, più "roba" prendo e dunque meglio medio: essendo rumore, per fare una caratterizzazione è necessario avere più potenza in ingresso all'elaborazione.

Capitolo 5

Caratterizzazione di dispositivi di potenza

Precedentemente si è già parlato di misure di potenza, quando si è parlato di power meters; a questo punto si vuole parlare ancora di misure di potenza, però più che altro focalizzandoci sulla caratterizzazione di dispositivi o sottosistemi di potenza, quali per esempio amplificatori di potenza.

Gli amplificatori di potenza, nelle classi più “tradizionali”, possono essere di tre tipi: A, B, AB, C; di queste, la più usata è senza dubbio la classe AB, che garantisce una buona efficienza e un buon guadagno. I parametri che permettono di scegliere la classe di amplificazione sono il **bias point** e la **retta di carico**. Il fatto di usare sistemi non lineari, tuttavia, genera armoniche, e questo porterà all’uso di carichi armonici, atti a cortocircuitare un certo numero di armoniche a massa. I carichi armonici avranno coefficienti di riflessione Γ con moduli molto elevati, e fasi che sostanzialmente dipendono dal transistor e dal package (ricordando che in un transistor non si ha solo la parte attiva, bensì anche la parte di parametri parassiti intrinseci, ossia propri del transistor come dispositivo, ed estrinseci, ossia dovuti al package). Tra i vari problemi che si hanno inoltre si ha l’effetto ginocchio: la zona di ginocchio de-idealizza il modello, introducendo una difficoltà su di un problema già complicato: nei modelli ideali si hanno transizioni istantanee tra regioni di non linearità e regioni di linearità; nella pratica, si ha un “ginocchio”, nel senso che la curva ha una transizione non brusca, la quale fa peggiorare la qualità della nostra caratterizzazione.

Misurare le forme d’onda della corrente dei transistori è estremamente complicato, ed è uno dei prodotti più avanzati e sofisticati che si riesca a realizzare oggi, in termini di strumentazione; la complicazione deriva dal fatto che tutte queste misure sono fortemente dipendenti dalla posizione alla quale si fa la misura stessa, dunque al fatto che esse saranno vincolate a un

piano di riferimento opportuno.

I problemi non sono solo questi: i transistori presentano diversi tipi di non idealità; la prima delle non idealità potrebbe essere la non linearità non solo del generatore, ma anche della porta di ingresso; spesso i dispositivi inoltre non possono neanche essere considerati reciproci, essendo $S_{12} \neq 0$.

5.1 Introduzione al load pull

Si è parlato precedentemente di curva di carico come di uno dei parametri che permettono di scegliere la classe di funzionamento del nostro sistema di amplificazione. Sarebbe cosa buona dunque avere dei mezzi in grado di permetterci di misurare la migliore retta di carico sotto il punto di vista di alcune caratteristiche del sistema costituito dal transistore e dalle varie reti di adattamento, in modo da ottenere il miglior compromesso possibile con gli altri parametri.

Si hanno diversi tipi di load pull: il load pull tradizionale, per cui si modifica la condizione di carico alla porta di uscita; questo può essere o alla fondamentale (quello più tradizionale), o alle armoniche, nel senso che si può scegliere di controllare le condizioni di carico sotto il punto di vista di armoniche superiori alla fondamentale. Cosa si fa sostanzialmente quando si fa load pull? L'idea è: prima di tutto si impone una certa condizione di carico, dunque per ciascuna di esse si effettua uno sweep di potenza, e si ottiene un certo insieme di curve:

Ciascun pallino rosso in realtà è composto da un certo insieme di curve, che son quelle che si misurano con lo sweep, a una certa condizione di carico.

Si parla di source pull oltre che di load pull, quando invece che la condizione di carico alla porta di uscita si controlla quella alla porta di ingresso.

Le curve che possono essere plottate sono dei tipi più disparati: la potenza di uscita al punto di compressione a 1 dB, la PAE, curve di intermodulazione, e si ottimizza nella seguente maniera: una volta fatto tutto ciò che è stato detto (definizione del carico e power sweep), si tracciano delle curve di livello, ossia qualcosa di concettualmente simile ai cerchi a guadagno costante dove però le curve finali non saranno cerchi bensì “patate”, dette “curve di Cripps”, ossia i luoghi dei punti tali per cui, sulla carta di Smith, si ha una certa quantità (P_{out} o PAE o altro che sia) costante; il fatto di aver a che fare con non linearità comporta la distorsione dei cerchi in queste curve, ma l'idea è sempre la stessa: identificando le curve a livello costante, si riesce a identificare il punto sulla carta di Smith, dunque il Γ , il carico, tale per cui si riesce ad ottimizzare una certa quantità.

Le curve di load pull, ossia le curve di Cripps, si potrebbero quasi predire a priori, mediante la teoria (per quanto essa sia molto complicata), a partire dalle curve DC; il fatto che i transistori tuttavia presentano elementi parassiti comporta problemi sotto questo punto di vista. In un LDMOS, uno dei dispositivi più interessanti da analizzare con questo tipo di tecniche (essendo dispositivi di potenza e dunque non lineari), quasi tutto è lineare (ovviamente, a parte il generatore di corrente); il package però continua a essere sconosciuto, quindi o si riesce a determinare una modellistica per il package, o non si può predire nulla con la teoria. Per fare la modellistica del package di solito quello che si fa è caratterizzare un die senza transistore (ossia il solo package) mediante i parametri scattering, dunque un solo transistore (senza package), quindi si identifica il modello.

Esistono banchi di load pull basati sostanzialmente su tre sistemi:

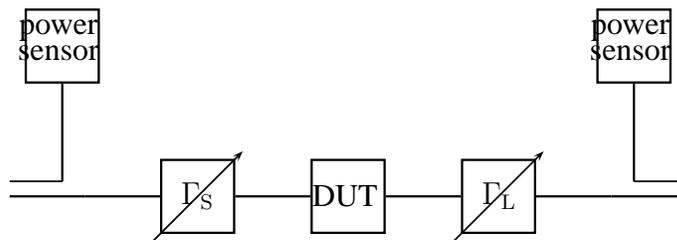
- basati sui power meter (o sui network analyzer scalari, che però non si utilizzano più);
- basati sui VNA;
- basati sulla misura delle forme d'onda nel dominio del tempo.

Nel mondo ci sono circa 2000 sistemi per effettuare misure di load pull; di questi, 50 sono realizzati mediante VNA, dieci o meno su sistemi time domain based, i restanti con power meters.

5.2 Load pull passivo

Verrà a questo punto analizzato un primo approccio per la realizzazione di un banco di lavoro per misure di load pull; il primo approccio che si può utilizzare è quello detto “load pull passivo”, ossia in cui si ha un carico variabile realizzato mediante soli elementi passivi. Lo schema che si considera è basato sull'uso di power meters.

L'idea sostanzialmente potrebbe essere quella di utilizzare uno schema di questo genere:



I tuner sono gli elementi che permettono di variare il Γ : si tratta in sostanza di elementi basati su di una *slab line*. Una slab line è una linea di trasmissione in cui si hanno due piani metallici, e un conduttore centrale; con un motore si cala una “carrier”, una “forchetta”, che può muoversi verso alto e basso, sinistra e destra. Queste possono essere realizzate mediante versioni manuali o elettromeccaniche. Lungo l’asse z la struttura si comporta come una linea di trasmissione, la cui impedenza caratteristica varia a seconda della posizione della carrier, fissata dal motore. Fino quando la distanza tra il centrale e la slag è sufficientemente grande, questa cosa si comporta come una coppia di linee di trasmissione che non parlano; se le avvicino, l’impedenza caratteristica della linea cambia, il carico è costante, dunque l’impedenza di ingresso vista cambia con la posizione della forchetta. Nella formula del calcolo del Γ si vede che Z_∞ cambia ma, essendo tutto reale, non si ha modifica della fase.

Quando la slag arriva molto vicina al centrale, tutto cambia drasticamente, dal momento che quando i conduttori si avvicinano le capacità di fringing si fanno sentire, la configurazione cambia, e ora non si ha più un modello semplice a linea di trasmissione, bensì un modello con delle capacità che cambiano e con Z_∞ che cambia.

Il fatto di avere il “bisciolino” comporta la necessità di avere una caratterizzazione puntuale del tuner, ossia una misura del tuner a ogni posizione, senza poter fare un modello. Questo, in altre parole, significa che il Γ_{in} non è modellabile in funzione delle posizioni verticali della forchetta, dal momento che, a causa di questo effetto delle capacità di fringing, non si riesce a predire in modo semplice il comportamento. Ci servono dei Γ elevati, dal momento che le impedenze di ingresso dei transistori sono basse (1Ω o qualcosa del genere per transistori di potenza). Lungo l’asse z si ha un normale spostamento di fase, dal momento che si aggiunge solo una linea di ritardo, ma quando si ragiona sull’asse verticale è tutto molto complicato.

5.2.1 Real-time load pull

Qual è il problema dei tuner? Essi sono elementi passivi, dunque sostanzialmente introducono perdite, e ciò ovviamente è molto negativo. Questo fatto ci spingerebbe verso il voler utilizzare una configurazione come la seguente:

I load modificabili sono i tuner; dal momento che siamo interessati a misurare il transistor, il DUT, la cosa migliore sarebbe misurare, dunque introdurre gli accoppiatori direzionali, direttamente sull’uscita del DUT, in modo da poter misurare direttamente i coefficienti di riflessione visti in questa posizione. Se non si facesse ciò, infatti, bisognerebbe conoscere molto bene i tuner, ossia precaratterizzarli, in modo da capire come essi modifichi-

no il Γ ; il tuner (si sta parlando sempre di sistemi passivi) è un sistema meccanico, dunque perchè precaratterizzare abbia un senso si deve avere anche una stabilità meccanica molto elevata del blocco. Si hanno poi, oltre ai problemi legati al tuner, i problemi legati al mondo esterno: se i connettori sono sbagliati, noi non possiamo accorgercene (se non osservandoli): l'errore potrebbe venire da un qualsiasi punto della catena, nella fattispecie dal DUT, punto che più frequentemente causa problemi.

Questo real time load pull sembrerebbe la soluzione a tutti i nostri problemi, ma in realtà a sua volta ha alcuni problemi. Poniamoci alcune domande, alle quali si risponderà man mano:

- Come si fa a misurare la potenza, con questo sistema?
- Qual è il massimo Γ che si riesce a far vedere al transistore? Nel dettaglio: se chiudo su un aperto l'uscita, ottenendo dopo il tuner $\Gamma_{\text{out}} = 1$, cosa si vede sulla porta di uscita del DUT?

Rispondere a quest'ultima domanda richiede l'uso della solita formula:

$$\Gamma_{\text{DUT,in}} = S_{11} + \frac{S_{12}S_{21}\Gamma_{\text{out}}}{1 - S_{22}\Gamma_{\text{out}}}$$

Gli S_{ij} che contano sono quelli degli accoppiatori direzionali e dei cavi che collegano le varie sezioni. Se dunque da un lato questo banco di misura real-time risolve il problema della misura del Γ , dal momento che prendiamo direttamente le onde di potenza dall'uscita del DUT, esso ha l'enorme limite di limitare la regione della carta di Smith nella quale è possibile fare l'indagine: dal momento che si passa per l'accoppiatore direzionale, si avranno sicuramente, anche avendo sull'uscita dell'accoppiatore $\Gamma_{\text{out}} = 1$, coefficienti di riflessione minori di 1 in modulo all'ingresso/uscita del DUT, a causa delle perdite di inserzione presenti negli accoppiatori direzionali, ottenendo dunque solo un sottoinsieme della carta di Smith.

Cerchiamo di capire quale può essere l'influenza degli accoppiatori direzionali in un sistema di questo tipo: immaginiamo di dover misurare un transistore di potenza, per esempio con 100 W in uscita, ossia 50 dB_m; questi devono essere misurati. Qual è un fattore di accoppiamento ragionevole per la gestione di una potenza così elevata? Beh, prima di tutto cerchiamo di identificare meglio il problema: se abbiamo 50dB_m all'interno di uno strumento esso sicuramente si romperà; la potenza che andrà misurata è quella accoppiata, dunque l'accoppiamento deve essere tale da prendere solo una piccola porzione di questo 50dB_m, in modo tale che i power meter o i suoi sensori non vengano danneggiati. Mettendo di voler entrare con 0 dB_m nello

strumento, potremmo usare un fattore di accoppiamento C pari a - 50 dB. Supponendo, con lo stesso sistema, di dover misurare un amplificatore con una potenza molto più ridotta, per esempio 20 dB_m, potremmo avere un fattore di accoppiamento di 25 dB.

Come detto, il problema sono le perdite di inserzione. Supponiamo di aver a che fare con un accoppiatore a 25 dB di accoppiamento; quali sono le perdite di inserzione? Beh, vediamo mediante un calcolo:

$$10^{-\frac{25}{20}} = 0,05623$$

questo numero può essere così utilizzato:

$$P_{\text{ins,loss}} = \sqrt{1 - (0,05623)^2} = 0,9984$$

Ora, se calcoliamo l'insertion loss in dB, si ottiene:

$$20 \log_{10}(0,9984) = -0,01376 \text{ dB}$$

Perdite di inserzione di 0,014 dB: assolutamente irrealistiche, tenendo conto che un accoppiatore commerciale come minimo ha perdite di inserzione di 0,2 dB. Immaginando di avere 0,2 dB di perdite di inserzione, chiudendo su un carico puramente reattivo (corto o aperto che sia), qual è il massimo Γ che posso portare al DUT?

$$\Gamma_{\text{DUT,max}} = 0 - 0,4 \text{ dB}$$

cosa significa? Vediamo:

$$10^{-\frac{0,4}{20}} = 0,955$$

Non è ancora male. Peccato che i couplers da usare di solito siano almeno due, per misurare onde incidenti e riflesse; supponendo di perdere 1 dB invece che solo 0,4 (si ricordi che è 0,4 e non 0,2 dal momento che il Γ va come e^{-j2kl} , si ha:

$$10^{-\frac{1}{20}} = 0,891$$

Considerando il segno negativo visto il carico iniziale, si ha:

$$\frac{1 - 0,891}{1 + 0,891} \times 50\Omega = 2,875\Omega$$

ossia si ha a che fare con una resistenza di circa 3 Ω .

Questo fatto degli accoppiatori per circa 20 anni è stato il collo di bottiglia per l'introduzione di un sistema real time di misura load pull.

5.2.2 Uso di wave probe

Usare accoppiatori direzionali per prelevare il segnale non è l'unica soluzione: esiste di fatto un'altra idea, basata sull'uso di probe, in grado di sentire il campo elettrico. Se si mette in un pezzo dove non c'è la slag una antenina, che farà da probe, si sente il campo elettrico; se si mette una spira, si sente il campo magnetico; i due campi sono tra di loro ortogonali, dal momento che la struttura che stiamo studiando è TEM, quindi, sentendo da una parte la tensione, dall'altra la corrente, si riesce a fare il sensing. In un certo senso questa struttura è direzionale, ma non nel senso classico, in cui si ha un accoppiatore che misura l'onda incidente e l'onda riflessa, ma, data la trasformazione tensione-onde, è direzionale, dal momento che il campo magnetico è ortogonale al campo elettrico. Come noto, le onde di potenza a e b sono ortogonali; se le probe di tensione e corrente misurano due grandezze ortogonali, allora si ha direzionalità ma in senso diverso rispetto a prima: invece che avere qualcosa che misura fisicamente a e b (l'accoppiatore direzionale), si ha una trasformazione che permette, a partire dalla misura di altre due grandezze, le quali saranno sempre ortogonali, di ottenere qualcosa di equivalente. Queste sono dette "wave probe" ma di fatto misurano altro, ossia tensione e corrente, dunque il nome non ha assolutamente senso. Lo svantaggio della cosa è che non si misurano le vere e proprie onde incidenti/riflesse, ma tensione e corrente.

Il modello di errore del VNA, che si utilizza per fare la misura, è assolutamente indipendente dal fatto che si abbia a che fare con un coupler o con wave probes: l'unica condizione che si era richiesta, al momento di proporre la teoria degli error box, era la linearità. Dal momento che dunque la calibrazione di questo oggetto è assolutamente analoga a quella del VNA, non è necessario chiederci se possiamo calibrare le wave probe, se non piuttosto quanto duri la calibrazione: la sensitivity rispetto al Γ che vogliamo trovare, delle wave probe, potrebbe essere decisamente peggiore di quella di un coupler.

Nel caso delle wave probe, l'error box sarà semplicemente data dalle trasformazioni modali:

$$v_1 = \sqrt{Z_0}(a_1 + b_1)$$

$$i_1 = \frac{1}{\sqrt{Z_0}}(a_1 - b_1)$$

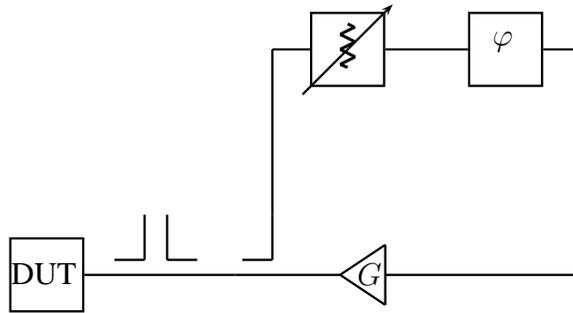
e così via.

Quando si parla di accuracy, si hanno delle differenze: gli accoppiatori direzionali più moderni, quelli inventati recentemente nel Politecnico

di Torino, sono di fatto idonei a misure a elevata accuratezza; le wave probe tendenzialmente no.

5.3 Load pull attivo

Finora è stato analizzato un banco di load pull passivo; questa non è l'unica possibilità, dal momento che è possibile allestire banchi in cui non serve avere un tuner. Un banco di load pull attivo potrebbe essere il seguente:



Si realizza un sistema di reazione: a partire da un accoppiatore, si prende parte dell'onda, la si manda verso "l'alto", la si attenua e la si rimanda indietro. Considerando l'onda riflessa a dal sistema e l'onda b incidente, si ha:

$$a_2 = b_2 C A e^{-j\varphi} G$$

dove C è il fattore di accoppiamento del loop coupler, A il fattore di attenuazione, G il fattore di guadagno, e φ all'argomento dell'esponenziale il termine di phase shift. Si può dunque dire che il Γ opposto da questo loop sia:

$$\Gamma = \frac{a_2}{b_2} = C A G e^{-j\varphi}$$

Invece che avere un carico passivo realizzato mediante un tuner, si ha un carico attivo: ciò permette di generare un coefficiente di riflessione dopo l'accoppiatore maggiore di 1, così che, una volta riportato indietro, si possa andare dove si vuole sulla carta di Smith.

Questa tecnologia ha moltissimi problemi, che so stati risolti 10 ÷ 12 anni fa.

Vi sono alcuni vantaggi e alcuni svantaggi; partiamo dagli svantaggi.

- L'amplificatore di loop deve essere estremamente lineare, perchè se esso introducesse armoniche potrebbe capitare di tutto. Si devono avere almeno 10 dB di potenza superiore alla massima potenza che vogliamo misurare; volendo però misurare un amplificatore da 100 W, ne servirebbe uno da 1 kW, e ciò non avrebbe senso, dunque si deve per forza usare con basse potenze. Dove si ha potenza, conviene evitarlo.
- Il carico attivo oscilla, o meglio una volta oscillava.

Ovviamente, vi sono anche dei vantaggi.

- Il carico attivo è molto più veloce di un carico elettromeccanico.
- Non necessita precaratterizzazione, essendo messo dopo l'accoppiatore.
- Può essere esteso alle armoniche in modo molto semplice.

Parliamo un po' del fatto che oscillava: per oscillare, un sistema con feedback deve avere modulo del guadagno di anello unitario. Il guadagno di anello è:

$$T = I \times A \times G \times e^{-j\varphi}$$

dove I è l'isolamento: si deve rimanere nell'anello. Supponendo che tutto sia adattato e si abbia un oggetto con delle perdite L , il massimo Γ che riusciamo a fare, Γ_{\max} , è:

$$\Gamma_{\max} = L^2 C A e^{-j\varphi} G$$

dove

$$L^2 = S_{12} S_{21}$$

La limitazione massima che possiamo avere per il guadagno di anello è:

$$T = I \times A \times G \times e^{-j\varphi} < 1$$

altrimenti se il guadagno di anello non soddisfa ciò abbiamo oscillazione. Sappiamo che la direttività ha un valore pari a:

$$D = \frac{I}{C}$$

A questo punto, l'unica incognita che rimane è il massimo valore di perdite che possiamo recuperare. Essendo il Γ_{\max} pari a 1, posso dire che:

$$1 = IAGe^{-j\varphi} \implies I = \frac{1}{AGe^{-j\varphi}} \triangleq \frac{1}{x}$$

dunque, possiamo dire che:

$$1 = L^2 C x$$

dunque, se ho D , posso dire che:

$$D = \frac{1}{x C}$$

dunque

$$C = \frac{1}{x D}$$

dunque

$$\Gamma_{\max} = \frac{L^2 \frac{1}{x D}}{\frac{1}{x}} = \frac{L^2}{D}$$

Questo è il risultato che ci interessava: la direttività deve essere maggiore del quadrato delle perdite. Questo significa che, se ho 1 dB di perdite, la direttività deve essere almeno maggiore di 2 dB. Se si ha però a che fare con cavi belli lunghi, con molte perdite, la direttività dovrà essere grande il doppio di queste perdite.

Problema: si supponga a questo punto di avere un sistema di load pull fatto così, con due accoppiatori direzionali, cavo, accoppiatore con 40 dB di direttività, in guida magari, il loop fatto come quello spiegato, perdite di circa 10 dB; se mettiamo un thru, tutto funziona, e si copre l'intera carta di Smith, avendo a disposizione un $\Gamma = 1$; nel caso però si metta un transistor, tutto oscilla, anche già con $\Gamma = 0,5$. Come mai? Come si può evitare?

Oscillazione significa che si ha un guadagno di anello unitario da qualche parte. Abbiamo visto che c'è un anello, ma in realtà non è l'unico: il fatto che con un transistor si abbia oscillazione e con un thru no è dovuto al fatto che l'anello che crea problemi è quello per cui si ha segnale che rimbalza tra il loop coupler e il measuring system. Questa cosa è causata dalla condizione di stabilità studiata sui parametri scattering: il fatto che il prodotto dei due Γ (vedi Elettronica delle Microonde) sia minore di 1. $\Gamma_{\text{out}}\Gamma_L$ è minore di 1, dal momento che non ho interesse di avere Γ_L fuori dalla carta di Smith, e Γ_{out} è dentro la carta di Smith a meno che il transistor non sia oscillante di suo (caso banale).

Si ha un Γ molto elevato, pari almeno al doppio delle perdite. Cosa fa dunque rendere instabile questo sistema? Quando vale il Γ visto dal sistema

di misura? Beh, trattando il blocco di misura come una matrice scattering, si avrà:

$$\Gamma_{\text{meas}} = S_{11}^{\text{meas}} + \frac{S_{12}^{\text{meas}} S_{21}^{\text{meas}} \Gamma_{\text{DUT}}}{1 - S_{22}^{\text{meas}} \Gamma_{\text{DUT}}}$$

Io voglio che il Γ proposto dal DUT sia unitario, e posso dire che esso vale:

$$S_{11} + \frac{L^2 \Gamma_{\text{DUT}}}{1 - S_{22} \Gamma_{\text{DUT}}} = S_{11} + \frac{L^2}{1 - S_{22}}$$

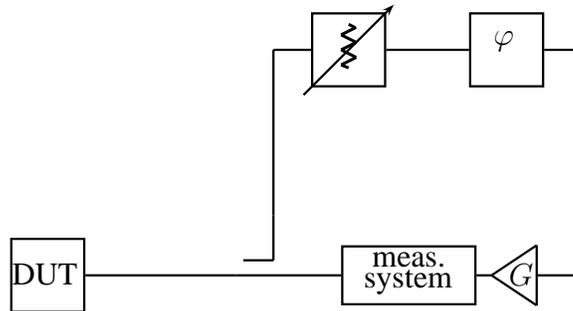
Questo è il Γ che si vede al measuring system, a partire dal DUT. Il Γ del measurement system dovrà essere questo, moltiplicato per $\frac{1}{L^2}$, in modo che esso possa essere maggiorato quanto basta per poter arrivare al DUT e valere 1. Dal momento che la condizione che ci interessa è il fatto che il prodotto dei due coefficienti di riflessione sia minore di uno, si avrà:

$$\left(S_{11} + \frac{L^2}{1 - S_{22}} \right) \frac{1}{L^2}$$

dal momento che $S_{22} \sim 0$, questa espressione diventa:

$$\frac{S_{11}}{L^2} + 1$$

A questo punto siamo riusciti a capire che cosa porta ad avere le oscillazioni: il disadattamento del sistema di misura, ossia il S_{11} . Questo problema si risolve con il seguente stratagemma:



Il modo di risolvere è prendere il measurement system e metterlo dentro all'anello, dopo il loop gain; in questo modo si usa il fatto di aver messo le perdite dentro e non fuori dall'anello, in modo che non si abbia più il rischio di oscillazione con Γ vicino a 1. L'unico effetto è che il guadagno complessivo dell'anello è quello di prima, con anche le perdite.

Nel loop si usa introdurre anche uno YIG filter: questo è utile dal momento che evita di far oscillare il loop interno, garantendo così la direttività su tutte le frequenze presenti nel sistema. Il sistema potrebbe infatti a oscillare a una frequenza che non è quella di misura, e lo YIG filter dunque la sopprime. Purtroppo questa cosa porta ad altre cose: cambiare la fase dello YIG significa sostanzialmente cambiare la fase del coefficiente di riflessione proposto al DUT; per questo motivo bisogna prestare attenzione al progetto del active loop. Sarebbe bello sostituire lo YIG filter, perchè esso è a banda veramente stretta; questo permetterebbe solo di fare misure in CW puro, ma al giorno d'oggi è necessario usare carichi a più larga banda.

Ciò che si usa fare ora è accoppiare un active loop con un tuner, inserito all'interno del loop; il Γ viene modificato sia in modulo sia in fase. Dal momento che si può considerare il sistema abbastanza lineare, è possibile applicare il principio della sovrapposizione degli effetti, e gli effetti sul Γ saranno tendenzialmente la sovrapposizione degli effetti del tuner e del active loop:

$$\Gamma_{\text{totale}} = \Gamma_{\text{tuner}} + \Gamma_{\text{loop}}$$

o comunque qualcosa del genere (col circolatore si ha un buon disaccoppiamento quindi quella formula è abbastanza vera). In realtà si tratta di una somma vettoriale, dunque dire che è la somma dei due significa avere questa situazione sulla carta di Smith:

Si possono far cambiare i due centri: modificare sia il coefficiente di riflessione del tuner sia quello del loop. Il tuner fa da pre-matching al loop, ossia riflette parte della potenza necessaria, a costo nullo. In questo modo si riescono a risparmiare fino a 7 dB di potenza di uscita dell'amplificatore di loop. Se si usa un accoppiatore direzionale con bassissime perdite si riesce a:

- fare da bootstrap al carico;
- si riesce a fare il tuning delle armoniche, impostando la frequenza dello YIG filter sulle armoniche invece che sulla fondamentale di misura.

Per l'intermodulazione di solito quello che si fa è agire sulla sorgente, dunque si fa un source pull: prima si può mettere a posto la sorgente, poi modificare l'uscita per massimizzare PAE o potenza di uscita.

5.3.1 Metodo Takayama per il load pull

Esiste un altro modo di realizzare il carico per il load pull, ossia il metodo Takayama: questo è basato sull'uso di tre sorgenti per ciascuna frequenza,

per ciascuna armonica. Queste sorgenti hanno la stessa base tempi e sono controllabili in fase, così che si riesca a controllare il modulo e la fase del segnale che ritorna. Si ha un isolatore, in modo che il segnale non vada indietro nelle sorgenti.

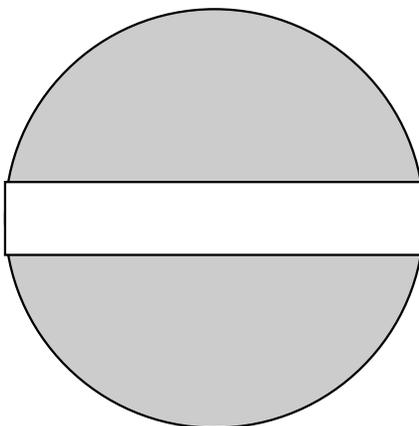
Il principio di questo carico attivo è: dato un generatore, un DUT, un altro generatore, un circolatore che isola l'isolatore, e si misura il b/a per determinare il Γ .

Il problema di questo oggetto è il fatto che idealmente si ha che:

$$b = \Gamma a$$

dove Γ è ciò che rappresenta il carico. Questo vuol dire che dal momento in cui io aumento a , il carico dovrebbe far aumentare anche b , ma di fatto non è detto che accada ciò: b infatti è fissato. Se si guardasse infatti sotto il punto di vista del dispositivo, per un certo a , si potrebbe fare in modo tale che si abbia un certo b e quindi ottenere una certa impedenza, ma questo non è come avere un carico: noi abbiamo a , abbiamo un certo b fisso, imposto, e quindi non si ha un vero carico ma qualcosa che solo gli assomiglia. Il problema sembrerebbe dunque solo di forma, ma non è così: come noto, un dispositivo è caratterizzabile mediante la sua curva $P_{in} - P_{out}$, la quale per un certo tratto è una curva lineare, ma dopo tende a saturare. Quando aumento dunque la potenza, il nostro "carico", invece di stare fisso nello stesso punto della carta di Smith, si muove verso il centro, dal momento che la potenza P_{out} inizia a saturare. Fino a quando il dispositivo è in linearità, costruire un algoritmo di correzione è abbastanza semplice; quando si esce dalla linearità, costruire questo algoritmo di correzione è molto più complicato, dal momento che, quando il "carico" si sposta, il dispositivo si comporta in modo diverso, dà una potenza diversa, e per questo continua a muoversi in modo imprevedibile. Un'idea fu quella di mappare tutto il comportamento, in modo da conoscere, data la posizione del "carico" in ogni punto, e gli spostamenti, il cambiamento del livello di potenza. Questo è un rischio ancora peggiore perchè, se non si fa attenzione, variando il carico si rischia addirittura di finire fuori dalla carta di Smith, dal momento che abbiamo cambiato il livello di potenza.

Al giorno d'oggi, si fa qualcosa di diverso: controllare in fase e in modulo i generatori è molto semplice, ed è estremamente veloce, agendo sul modulatore I/Q. Questa cosa è però pericolosa perchè per raggiungere un alto valore di Γ siamo costretti ad andare vicino alla carta di Smith, e se si aumenta la potenza il carico rischia di andare fuori da essa. Il rischio deriva dal fatto che i transistori di potenza potrebbero avere delle regioni di instabilità molto particolari, per esempio di questo tipo:



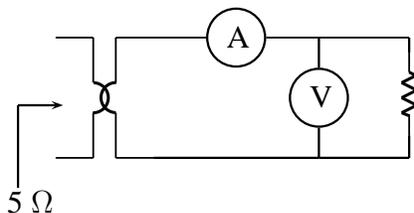
Fare uno sweep in automatico esaustivo della carta di Smith rischia di far entrare il dispositivo nella regione di funzionamento instabile, dal momento che essa ha questa forma molto particolare, portando dunque il dispositivo a bruciarsi.

Ciò che però si può fare con questo sistema è il “waveform engineering”: si può sintetizzare perfettamente fino alla terza armonica dell’onda riflessa, dunque il carico che vogliamo, e usando un oscilloscopio sampling si può modificare la forma d’onda ai piani di riferimento del dispositivo. Questo permette di visualizzare le forme d’onda time-domain. Si tornerà tra breve sulla cosa.

5.3.2 Fixtures

Si vuole a questo punto introdurre un certo insieme di sottosezioni, in modo tale da completare alcuni discorsi.

Molto importante è di sicuro il discorso delle fixtures. Si è visto che i limiti dei carichi passivi sono legati al fatto che il coefficiente di riflessione proposto dal tuner è limitato dalle perdite del test set. Una delle soluzioni è dunque quella di usare un trasformatore:



Tutto il nostro sistema di misura funziona a 50 Ω, mentre il transistoro funziona a 5Ω o meno: usando un sistema di questo genere dunque si può far

vedere al transistore quello che si vuole, e misurare alla impedenza normale. Tutto ciò sembra facile, ma fare il trasformatore è di fatto complicato.

Per un certo periodo si è fatto uso dei pre-matched network:

Facendo così si ha un problema di calibrazione che non finisce più: bisognerebbe calibrare sia il tuner sia il pre-match tuner. Ciò che alla fine si è fatto è questo:

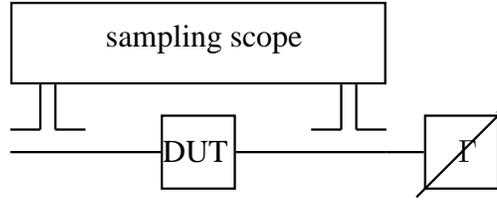
Usare delle reti di adattamento. Si usano dunque dei trasformatori di questo tipo. Può però succedere che si vada in una zona di instabilità. Con l'active loop è più difficile che capiti ciò, dal momento che il filtro YIG è a banda piuttosto stretta. Si ha inoltre il problema della dipendenza dalla frequenza: a 1,8 GHz e a 2,7 GHz per esempio le impedenze mostrate sono completamente diverse; se si usassero per esempio due tuner diversi, senza aver fatto loro la caratterizzazione a ogni frequenza che si intende utilizzare, non si ha idea di quale impedenza essi proponano. Con il sistema real time si sa che le impedenze sono diverse, ma le si misura; usando passive tuner, in altre parole, non si conoscono le condizioni alle armoniche. Ciò che si suggerisce di fare, nel caso si usi un passive tuner, dunque, è mettersi nella condizione importante, tunare, togliere il dispositivo e sostituirlo con un thru, ri-misurare la condizione di carico ottimo, e così si è sicuri che la condizione sia quella giusta.

5.4 Time-domain analysis: Non-linear Network Analyzer

Si vuole a questo punto presentare l'ultima tecnica di misura per i dispositivi di potenza, ossia quella della visualizzazione delle forme d'onda nel dominio del tempo. Si tratta, come in molti altri casi (quale per esempio quello degli analizzatori di spettro), di misura ottime sotto il punto di vista qualitativo, nel senso che esse permettono di comprendere il comportamento del dispositivo, ma sotto il punto di vista quantitativo non si parla assolutamente di misure valide.

Gli strumenti che fanno questo tipo di misure sono i cosiddetti "nonlinear VNA", dal momento che sono comunque strumenti in grado di misurare dei parametri scattering, ma in questo caso non lineari (si pensi al PNA-X, che misura gli X parameters). Per studiare questo tipo di dispositivi si possono riprendere molti concetti noti dal VNA.

Lo schema di principio di un nonlinear VNA potrebbe essere il seguente:



Questo oggetto va calibrato: dato un tuner, ossia un Γ variabile, è necessario vedere Γ_{in} , Γ_L , P_{in} , P_{out} . Al fine di fare la calibratura, ci servono a_1 , b_1 , a_2 , b_2 .

Vediamo: nel sampling scope ho delle misure di tensioni variabili nel dominio del tempo. La procedura è quella di prendere le onde misurate nel dominio del tempo, $a_m(t)$ (che si può vedere come un vettore, dal momento che la caratterizzazione nel dominio del tempo richiede comunque l'uso di un certo numero di campioni dell'onda nel dominio del tempo, dunque comunque una n -pla), e passare ai loro fasori, $a_m(f)$, dunque fare la stessa cosa con le onde. Si ricordi che abbiamo dei grossi limiti quando vogliamo usare la FFT: essa dovrebbe, come già visto in precedenza, richiedere una stima spettrale, cosa molto complicata da fare; dal momento che noi tuttavia sappiamo a che frequenza stiamo misurando e tendenzialmente anche quali armoniche si vengono a formare, possiamo evitare la stima spettrale.

Quello che si fa è un qualcosa di concettualmente simile a quanto si studia come approccio per il procedimento di analisi a regime, "harmonic balance": si ha una certa matrice opportuna che, se applicata ai vettori, permette di passare da campioni nel dominio del tempo a campioni nel dominio della frequenza. La funzione nel tempo sarà rappresentata sotto forma di n -pla, come già detto. Si ha qualcosa del tipo:

$$\underline{\underline{M}} \underline{a}(t) = \underline{a}(f)$$

$\underline{\underline{M}}$ è una matrice di coseni e seni, ossia una matrice che permette di passare dal dominio del tempo al dominio dei fasori; tutti i vari coseni avranno pulsazione già nota, dal momento che sostanzialmente si ha o la fondamentale o delle sue armoniche:

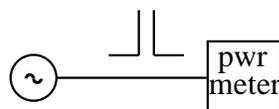
$$\underline{a}(t) = \sum_i a_i \cos(\omega_i t)$$

Una volta date le misure è possibile applicare il solito modello 4-sampler, e determinare le onde reali a partire da quelle misurate, passando per l'error box $\underline{\underline{E}}$:

$$\begin{bmatrix} \underline{a}_1 \\ \underline{b}_1 \end{bmatrix} = \underline{\underline{E}} \begin{bmatrix} \underline{a}_{m1} \\ \underline{b}_{m2} \end{bmatrix}$$

Da tutta la teoria che abbiamo analizzato precedentemente, \underline{E} è nota a meno di una costante; rispetto a quanto fatto precedentemente, però, si ha qualcosa di diverso: in tutti i ragionamenti fatti prima, si ragionava su misure di rapporti, dunque non si aveva la necessità di conoscere tutti i parametri (8), bensì uno in meno, a partire dal quale si potevano definire tutte le matrici; ora avremo 8 coefficienti da conoscere e non 7, dal momento che siamo interessati alle onde e non più ai rapporti (quale per esempio era il coefficiente di riflessione).

Ripartiamo dunque dal caso a 1 porta:



Dalla taratura è possibile trovare e_{00} , e_{11} , $e_{10}e_{01}$ (il prodotto degli altri). Con un power meter alla porta 1 dunque si può misurare il modulo di e_{10} , e a meno della fase posso dunque conoscere tutti i moduli dei vari parametri dell'error box. La taratura di potenza si aggiunge semplicemente aggiungendo alla taratura normale la misura di un power meter a una delle porte. Si noti che noi siamo nel dominio della frequenza, e stiamo facendo tutta la nostra elaborazione nel dominio della frequenza.

Per ritornare nel dominio del tempo ci serve l'informazione di fase, ma non solo: la questione è assai complicata, dal momento che quando noi scriviamo

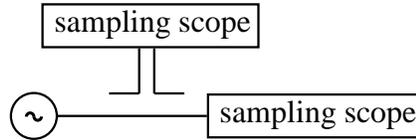
$$a(t) = \sum_i a_i \cos(\omega_i t + \varphi_i)$$

di fatto ci serve avere φ_i , e ciò è estremamente problematico: le misure che facciamo sono frequenza per frequenza, ossia, data una certa frequenza f_0 , si fan le varie misure e i calcoli, ottenendo e_{ij} alla suddetta frequenza; metto f_1 , rifaccio la stessa cosa, e così via. Per costruire il fasore, però la φ_i deve essere costruita rispetto allo stesso istante di tempo: tutti i fasori funzionano dal momento che la dipendenza dal tempo per ciascuno degli oggetti in cui abbiamo scomposto, è sempre la stessa: sempre la stessa base tempi. Il problema è che noi non sappiamo, quando si fa l'analisi frequenza per frequenza, quale sia il riferimento di fase comune, ma, dal momento che queste misure si fanno con dei PLL, noi sappiamo che si ha una differenza di fase costante, ma quanto essa valga è un mistero.

L'obiettivo dunque del momento è quello di ottenere in qualche modo un riferimento di fase comune alle varie misure.

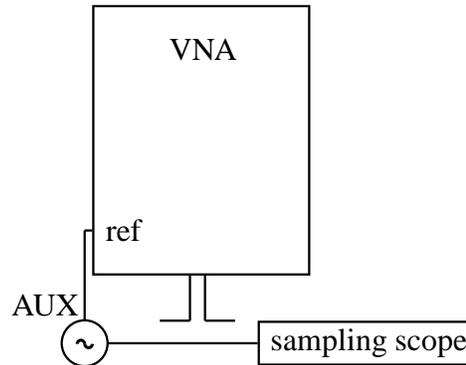
Il problema elementare con cui abbiamo a che fare è: lavorando a un certo piano di riferimento, si ha un certo sampling scope, il quale può produrre in

uscita una forma d'onda nel dominio del tempo (a differenza dell'analizzatore di reti che produce fasori, con una fase però scoerente, dal momento che il PLL si aggancia come vuole). L'idea è stata quella di uniformare le due cose:



Per misurare e_{10} basta usare un altro sampling scope, al posto del power meter, misurare la forma d'onda in quel punto (mettendo ovviamente la stessa base tempi), dunque misurare la differenza di fase tra i due strumenti, $\Delta\varphi$: questa sarà la fase del coefficiente e_{10} . L'unica misura che faccio è quella di fase, come differenza di fase dei due. Il sampling scope in uscita dà già prodotti coerenti, informazioni coerenti tra di loro.

Urs Lott, al Politecnico di Torino, risolse la cosa utilizzando direttamente gli analizzatori di reti: prese un canale in più dell'analizzatore di reti, prese una sorgente alla fondamentale, e un diodo step recovery al reference (aggiunto in seguito). Senza il diodo, si può prendere la sorgente alla fondamentale, uscita ausiliaria (per esempio nel lavoro iniziale l'uscita AUX del 8340). Mandando la fondamentale da una parte (reference, dal AUX) e la moltiplicata dall'altra, stesso generatore, si riesce ad avere l'armonica coerente, dunque fasori coerenti, e questo per tutte le armoniche. I canali possono essere aggiunti al VNA mediante switching box, dunque si possono fare i vari $V_{\text{test}}/V_{\text{reference}}$ e trovare dei set di fasori coerenti.



La differenza tra l'usare un sampling scope e un VNA è il fatto che il VNA ha un elevatissimo dynamic range: il filtro IF finale del VNA infatti ha banda molto più stretta, mentre quello dei sampling scope ai MHz: il sampling scope ha molti meno bit equivalenti, oltre ad essere di suo meno accurato: tira fuori molto più rumore.

Si noti un fatto: con la soluzione VNA si hanno per ora fasori coerenti tra loro, dunque riferiti alla stessa base tempi, ma non “assoluti”: non si ha un riferimento rispetto al quale si valuta la fase, nel senso che si sa che si ha una relazione di fase ben definita tra i vari fasori, ma non si sa quale sia la relazione tra questa fase coerente e la fase di riferimento. La soluzione per mettere a posto ciò si può risolvere mediante un sampling scope: introducendo alla porta interessata un sampling scope, è possibile sincronizzare le fasi. Purtroppo questa sincronizzazione avrà accuratezza che dipende dal sampling scope, che sarà un collo di bottiglia sotto questo punto di vista.

Non è tutto: qua stiamo ragionando sulle fasi, e quando si parla di fasi è obbligatorio fare molta attenzione ai piani di riferimento; in qualche modo il sampling scope va collegato, ma dunque servono dei fili, che introdurranno dello sfasamento. Il sampling scope campiona quello che gli entra dentro, dunque prende il segnale e lo shifta a frequenze più basse. Il cavo di collegamento di fatto fa sfasare e_{ij} , ma tutti di uno stesso valore, dal momento che le misure passano tutte per lo stesso filo, dunque si ha sì uno sfasamento, ma uguale per ogni valore, a meno che il filo non introduca dispersione (e dunque modifichi in modo diverso ciascuno dei valori).

La soluzione è avere un banco “ottimo”, calibrare un diodo detto “golden diode” che con in ingresso una certa frequenza produca esattamente le armoniche, e dunque utilizzare questo come riferimento preciso, per la calibrazione dell’analizzatore di reti non lineare. Una volta precaratterizzato il golden diode è possibile effettuare la calibrazione. Urs Lott invece che avere il diodo fisico realizzò un modello SPICE di questo, e su questo calibrò l’analizzatore di reti. Resta ovviamente il problema dell’accuratezza del sampling scope. Il consiglio è quello di tarare con un sampling scope, supponendo la linearità dei mixer.

Una volta costruito il nonlinear network analyzer, si vuole fare del load pull con esso! Studiare dunque la variazione del carico nel dominio del tempo (load-pull time domain): un banco in grado di misurare le forme d’onda nel dominio del tempo al variare delle condizioni di eccitazione e di carico.

Si possono tirare fuori informazioni interessanti; un esempio riguarda la PAE, ossia il fatto che si riesce a incrementare di un 10 % la PAE al massimo, al variare del carico armonico.