Curva del diodo e fit con gnuplot

fuso@df.unipi.it; http://www.df.unipi.it/∼fuso/dida

(Dated: version 1 - FF, 25 febbraio 2014)

Questa breve nota tratta della mia esperienza personale nel provare a fare un best-fit dei dati che ho personalmente raccolto nell'esperienza della curva caratteristica del diodo. Ho sottolineato il carattere personale dell'esperienza perché non è affatto detto che i miei risultati siano generali, sia perché i dati e le condizioni sperimentali potrebbero essere diverse da quelle che avete incontrato voi, sia perché la procedura di best-fit, essendo numerica, non è necessariamente replicabile quando i dati inseriti sono diversi.

I. INTRODUZIONE

Eseguire un best-fit con un metodo numerico, come quello che adotta il software gnuplot, è sempre una procedura delicata e che va tenuta sempre sotto il massimo controllo. Questo è particolarmente vero quando la funzione suggerita dal modello è "ostica" dal punto di vista numerico, come succede con le funzioni esponenziali, trattando le quali il computer si trova a registrare brusche variazioni in corrispondenza di piccole variazioni dell'argomento. Inoltre la possibilità di avere parametri fortemente correlati tra loro contribuisce non poco a rendere la procedura di best-fit e la sua corretta convergenza un terno al lotto. Quello che racconto qui è storia vera, basata su dati veri (presi un po' alla carlona, cioè di fretta) e su un trattamento vero con il mio piccolo computer (con Windoze installato...).

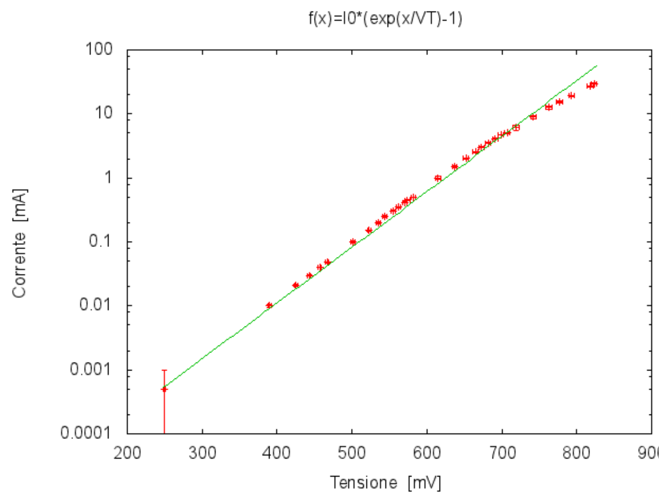

Figura 1. Dati e best fit secondo la funzione di Eq. 1.

II. DATI E MODELLO

I dati sono quelli della corrente che passa attraverso un diodo a giunzione bipolare di silicio (il modello dovrebbe essere l'1N914) al variare della tensione di polarizzazione (diretta). Per gli errori ho preso il più grande tra l'errore di misura (1 digit per il multimetro digitale usato come voltmetro e mezza tacchetta per quello analogico usato come amperometro) e l'accuratezza dichiarata per lo strumento. Si potrebbe opinare che l'accuratezza sovrastimi l'errore, in quanto fondamentalmente errore di calibrazione, però la necessità di cambiare scala per coprire l'intero intervallo di misure consiglia fortemente di considerare (anche) l'accuratezza come sorgente di errore. Per risparmiare nell'inserire i dati nel computer ho scelto di misurare le tensioni in mV e le correnti in mA (notate che il loro rapporto è ohm!).

Il modello di riferimento è quello della *giunzione bipolare di Shockley*:

$$I = I_0(\exp(\frac{V}{\eta V_T}) - 1) , \qquad (1)$$

dove I_0 rappresenta la *corrente di saturazione inversa*, attesa dell'ordine del nA, o decine di nA, $V_T = k_B T/e$ è la differenza di potenziale dovuta alla temperatura (ricordate: $k_B T \approx 1/40$ eV a temperatura ambiente) e η è un parametro che dipende dalle caratteristiche della giunzione e dovrebbe valere circa 2 nel caso di un diodo al silicio. A conti fatti, ci si attende $\eta V_T \approx 50$ mV.

La Fig. 1 mostra i dati, in scala semilogaritmica, assieme al risultato del best-fit secondo la funzione di Eq. 1 si vede subito a occhio che il risultato è pessimo. In effetti gnuplot si rifiuta di far variare il parametro ηV_T (posto pari a 50 mV come valore iniziale) e dunque esegue il fit cambiando, fino a quella che lui presume essere la convergenza, il parametro I_0 (che risulta 3.2 nA). In queste condizioni si ottiene un chi-quadro ridotto gigantesco (oltre 200!), ma non escludo che, imponendo un diverso valore iniziale per ηV_T, questo possa essere ridotto "a mano". Ci sono vari possibili motivi per spiegare la defaillance: secondo me il principale è la forte correlazione che esiste tra i due parametri del fit (per gnuplot è oltre 0.98).

In effetti nella regione di maggiore interesse, cioè per $V > V_{thr}$, che è la soglia del diodo, da dove la corrente aumenta sensibilmente, la funzione può essere approssimata cancellando il -1. Dunque i due parametri sono davvero molto correlati: gnuplot ritiene che variando ηV_T non ci siano apprezzabili variazioni del chi-quadro e si accontenta di variare solo I_0, con il risultato che potete vedere.

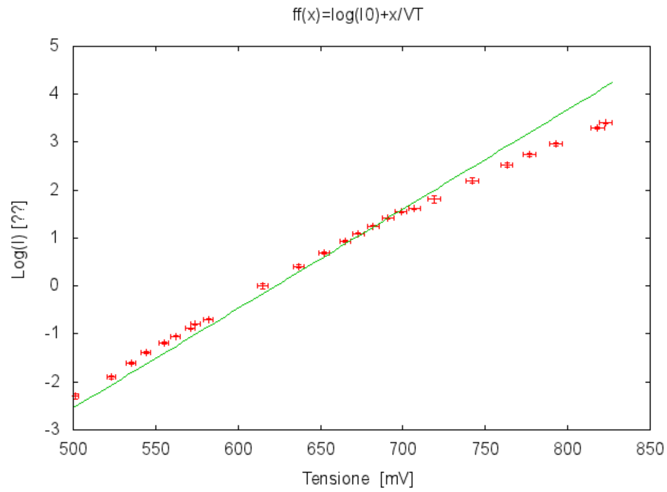

Figura 2. Dati e best fit secondo la funzione di Eq. 2: leggete il testo, riporta un sacco di osservazioni rilevanti!

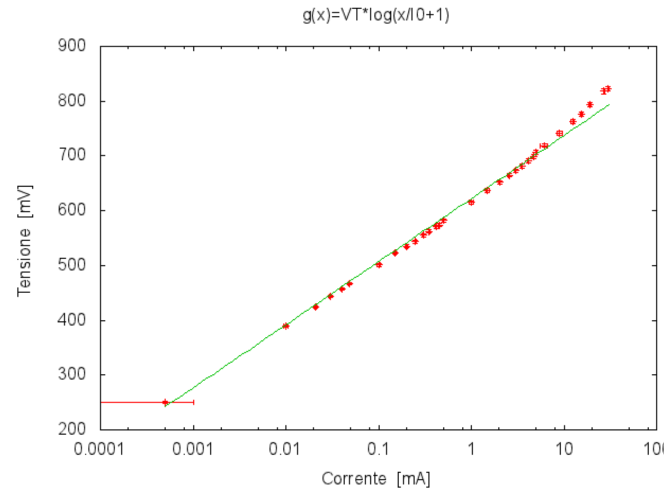

Figura 3. Dati e best fit secondo la funzione di Eq. 3.

Ci sono diverse possibilità per migliorare, almeno virtualmente, la situazione. Una prima possibilità consiste nel togliere il -1 della funzione e nel farne il logaritmo dei due membri. Si ottiene

$$\log(I) = \log(I_0) + V/(\eta V_T) \ . \qquad (2)$$

Questa è l'equazione di una retta e i due parametri perdono la correlazione che avevano prima.

In questo caso occorre graficare e considerare per il fit i dati di $\log(I)$. Ovviamente non c'è alcun bisogno di costruire una tabella con il logaritmo, dato che possiamo, come ben sapete, istruire gnuplot in modo opportuno. In sostanza per il grafico (che stavolta deve essere fatto in scala non logaritmica!) l'istruzione è `plot 'diodo.txt' using 1:(log($2)):3:($4/$2) w xyerrorbars`, per il fit `fit ff(x) 'diodo.txt' using 1:(log($2)):($4/$2) via I0,VT`, dove la funzione è `ff(x)=log(I0)+x/VT` (si intende, qui e nei titoli dei grafici, che il parametro di fit è in realtà ηV_T; naturalmente non è possibile distinguere con il fit tra due parametri che si moltiplicano tra loro). Come notate da queste istruzioni, mi sono anche preoccupato di propagare in modo corretto l'incertezza (controllate!); inoltre, per tenere conto dell'approssimazione sull'esponenziale citata prima, ho anche ristretto l'intervallo di analisi a tensioni superiori a 500 mV.

Il risultato è riportato in Fig. 2. Non è ancora affatto soddisfacente. Intanto c'è un brutto problema: non so cosa mettere come unità di misura dell'asse verticale (il logaritmo naturale di una corrente non è un'unità fisica!). Poi il fit non è proprio buono e c'è ancora un chi-quadro ridotto gigantesco.

A. Un modo più elegante

Esiste un modo decisamente più elegante e corretto per ottenere un risultato migliore. Esso consiste nel lasciar stare il -1 della funzione, e nell'invertirla, cioè nell'esprimere V in funzione di I. Si ottiene facilmente

$$V = \eta V_T \log(I/I_0 + 1) \ . \qquad (3)$$

Il risultato è mostrato in Fig. 4: notate che gli assi sono invertiti rispetto a prima, e la rappresentaione ora è logaritmica per l'asse orizzontale. Inoltre stavolta non incontro problemi nel definire le unità di misura (rappresento i dati in scala logaritmica, e non il logaritmo dei dati!). Però non posso sicuramente dirmi ancora soddisfatto: il chi-quadro ridotto scende a circa 10, la correlazione tra i parametri di fit scende a 0.96, ma l'accordo ancora non è molto buono e di conseguenza non mi fido troppo dei parametri ottenuti.

B. La fisica

L'ultima carta che decido di giocarmi non ha a che fare con la matematica, cioè l'impostazione della procedura di best-fit, ma invece riguarda la fisica (che notoriamente è più potente della matematica!). C'è infati un aspetto che il modello di Shockley trascura: il diodo che abbiamo usato ha piccole dimensioni e quindi è ragionevole supporre che il contatto tra giunzione e reofori (i filini) presenti una resistenza non trascurabile. Questo contatto può essere ritenuto ohmico, almeno approssimativamente, e per un resistore ohmico, anche fatto di un buon conduttore (alluminio, rame, o qualche lega metallica) la resistenza diventa non trascurabile quando la sezione è piccola, come è probabilmente nel nostro caso.

Decido allora di aggiungere un termine ohmico all'Eq. 3 (la resistenza del contatto è ovviamente in serie

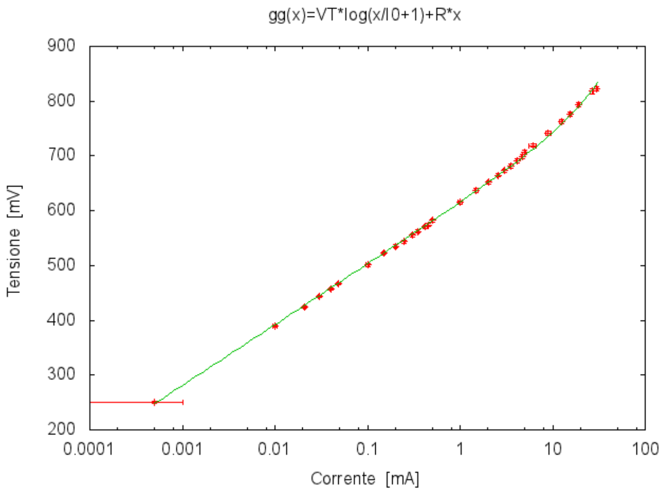

Figura 4. Dati e best fit secondo la funzione di Eq. 4.

alla giunzione), ottenendo

$$V = \eta V_T \log(I/I_0 + 1) + RI \ , \qquad (4)$$

dove R è un ulteriore parametro del fit che ha le dimensioni di una resistenza.

Il risultato stavolta è convincente: ottengo un chi-quadro ridotto di circa 1, e i valori ottenuti dal fit sono "ragionevoli": $I_0 = (2.9 \pm 0.1)$ nA, $\eta V_T = (48.2 \pm 0.2)$ mV, $R = (1.8 \pm 0.1)$ ohm (naturalmente mi sono preoccupato di dividere il risultato bruto dell'incertezza sui parametri data da gnuplot per l'`rms of residuals`). Inoltre la correlazione massima tra i parametri è scesa a 0.92 e posso ragionevolmente modellare l'intero range di tensioni/correnti esplorato. Bene, chi la dura la vince!

Come commento finale, ricordo che i problemi che ho incontrato e cercato di risolvere sono personali, cioè successi a me. Non è detto che per un altro set di dati, un altro computer, un'altra scelta dei parametri iniziali del fit si incontrino le stesse difficoltà che ho trovato io. Tuttavia il messaggio che impone di usare attenzione quando si esegue un best-fit deve rimanervi ben chiaro!