Computer Physics Communications 207 (2016) 193–201

Contents lists available at ScienceDirect

Computer Physics Communications

journal homepage: www.elsevier.com/locate/cpc

# On the very accurate numerical evaluation of the Generalized Fermi–Dirac Integrals

N. Mohankumar [a,*], A. Natarajan [b]

[a] 40, DJ Nagar, Peelamedu, 641004, India

[b] 26, BBC City Park, II Phase, Chennai, 600116, India

ARTICLE INFO

*Article history:*
Received 23 November 2015
Received in revised form
7 June 2016
Accepted 10 June 2016
Available online 22 June 2016

*Keywords:*
DE
Trapezoidal and Gauss–Legendre quadratures
Residue correction
Continued fraction algorithm

ABSTRACT

We indicate a new and a very accurate algorithm for the evaluation of the Generalized Fermi–Dirac Integral with a relative error less than $10^{-20}$. The method involves Double Exponential, Trapezoidal and Gauss–Legendre quadratures. For the residue correction of the Gauss–Legendre scheme, a simple and precise continued fraction algorithm is used.



## 1. Introduction

The Fermi–Dirac integral (FDI)

$$F_k(\eta) = \frac{1}{\Gamma(1+k)} \int_0^\infty \frac{t^k\, dt}{e^{t-\eta}+1} \tag{1}$$

is needed in a variety of problems involving the Fermi–Dirac distribution like the calculation of charge densities of semiconductor devices. A detailed account is given by Blakemore [1] for the various exact and approximate expressions available in the literature with special emphasis on $F_{1/2}(\eta)$. If the upper limit of integration is finite, the above integral is called the incomplete FDI. Goano [2] provides a large and accurate collection of algorithms to evaluate the ordinary as well as the incomplete FDI. The work in this area broadly consists of two groups. The first set deals with series expansions that are valid for small values of $\eta$ and the asymptotic approximations which are valid for large values of $\eta$ [3–10]. The second set consists of numerical algorithms, based either on rational approximation [11–13] that combine both high accuracy and minimum effort or they rely on numerical integration [14].

An integral related to the FDI that is needed in astrophysical problems like the stellar evolution is defined by

$$F_k(\eta,\theta) = \int_0^\infty \frac{t^k \sqrt{1+\theta\, t/2}\, dt}{e^{t-\eta}+1}. \tag{2}$$

This is the Generalized Fermi–Dirac Integral (GFDI) that depends on three parameters $k$, $\eta$ and $\theta$. When the parameter $\theta$ is zero, the GFDI reduces to the FDI without its gamma function term in the denominator.

As we remarked earlier, a detailed review of the asymptotic and the series expansions for the FDI is available in Blakemore [1]. For the sake of completeness, here we outline very briefly some of these approaches. Many of these expressions are derived from the classical series expansions provided by McDougall and Stoner [6] and Dingle [4,5]. The following is a typical one valid for integer and half-integer values [3].

$$F_k(\eta) = \sum_{r=1}^{\infty} (-1)^{r+1} e^{r\eta}\, r^{-(k+1)}; \quad \eta \ll 0$$

$$F_k(\eta) = \cos(k\pi) F_k(-\eta) + \frac{\eta^{k+1}}{\Gamma(k+2)}[1+R_k(\eta)]; \quad \eta > 0$$

$$R_k(\eta) = \sum_{r=1}^{\infty} \frac{\alpha_r\, \Gamma(k+2)}{\eta^{2r}\, \Gamma(k+2-2r)}; \quad \alpha_r = \sum_{n=1}^{\infty} (-1)^{n+1}\, 2\, n^{-2r}.$$

For evaluating the FDI, Goano [7] utilizes the fact that the term $1/[e^{t-\eta}+1]$ can be expanded in a geometric series for the cases

* Corresponding author.
*E-mail addresses:* kovainmk@gmail.com (N. Mohankumar), anat1946@yahoo.co.in (A. Natarajan).

http://dx.doi.org/10.1016/j.cpc.2016.06.004
0010-4655/

$t > \eta$ and the $t < \eta$ separately. The resulting series is integrated term by term and the final quantities are expressed in terms of the Kummer confluent hypergeometric functions of the first and second kind, $M(a, b, z)$ and $U(a, b, z)$, respectively, for which efficient algorithms exist.

$$F_k(\eta) = \frac{\eta^{k+1}}{\Gamma(k+2)}\left\{1 + \sum_{n=1}^{\infty}(-1)^{n-1}[(k+1)\,U(1, k+2, n\eta) - M(1, k+2, -n\eta)]\right\}.$$

Bhagat et al. [13] derive series expansion for the FDI. The approach here is to expand the term $1/[e^{t-\eta}+1]$ in a series after appropriate manipulations and then integrate the resulting integrand term by term. Next, to expedite convergence, acceleration technique like the Levin transform is made use of. It must be noted that this approach works better if $\eta$ is *large*. Again, for the GFDI, the term $\sqrt{1+\theta t/2}$ is also expanded in a series (possible *only for small values of* $\theta$ like $10^{-3}$) and the resulting series is integrated term by term and this is followed by convergence acceleration.

The FDI can be expressed in terms of the Zeta or the Polylogarithm or the incomplete Gamma functions. Below, we indicate an *exact series expression* for $F_k(\eta)$ in terms of the incomplete Gamma function [15]. We note the Mittag-Leffler expansion of sech$(z)$ function [16] which leads to the following modification of the denominator of the integrand of the FDI.

$$\mathrm{sech}(z) = \frac{1}{\cosh(z)} = \pi\sum_{l=0}^{\infty}\frac{(-1)^l\,(2l+1)}{z^2 + [(2l+1)\pi/2]^2}$$

$$\frac{1}{e^{t-\eta}+1} = \frac{e^{-(t-\eta)/2}}{2\,\cosh[(t-\eta)/2]}.$$

With the substitution above, we get the FDI as

$$a_l = \eta + i(2l+1)\pi; \qquad \phi_l = \tan^{-1}[(2l+1)\pi/\eta];$$
$$l = 0, 1, 2, \ldots$$

$$F_k(\eta) = \frac{2\pi\,e^{\eta/2}}{\Gamma(k+1)}\sum_{l=0}^{\infty}(-1)^l\,(2l+1)\int_0^{\infty}\frac{t^k\,e^{-t/2}\,dt}{(t-\eta)^2 + [(2l+1)\pi]^2}$$

$$F_k(\eta) = 2\,\mathrm{Im}\left\{\sum_{l=0}^{\infty}e^{i[\pi(k-1/2)-k\phi_l]}\,[\eta^2 + ((2l+1)\pi)^2]^{k/2}\,\Gamma(-k, -a_l^*/2)\right\}; \quad k > (-1).$$

After this brief outline of the series expansions, we turn to the numerical evaluations based on quadrature. Both Pichon and Sagar [17,18] use the modified Gauss–Laguerre schemes to achieve a better accuracy. Gautschi also constructs modified Gaussian schemes [19]. But these three methods [17–19] need a lot of computational effort since the weight and the node generation is a non-trivial task and also the weights and nodes change with the parameters $k$, $\eta$ and $\theta$. The GFDI and its derivative with respect to its parameters were evaluated by Gong et al. [20] by splitting $(0, \infty)$ into four intervals. The Gauss–Legendre scheme is used in the first three intervals and the Gauss–Laguerre scheme is used in the last interval. Here, the choice of break points is by trial and error.

The convergence of any quadrature scheme for the evaluation of the integrals defined by Eqs. (1,2) is impaired by the singularities of the integrands. If $k$ takes half-integer values (as in the case of the semiconductor device modeling) like $k = -(1/2), (1/2), (3/2), (5/2), \ldots$, then the origin $t = 0$ is a branch point. In addition, the integrands of both the GFDI and the FDI have a countable infinity of simple poles at $t_j$ defined by $t_j = \eta + i(2j+1)\pi$; $j = 0, \pm1, \pm2, \ldots$. When $k$ is a half-integer, the branch point singularity can be removed by setting $t = x^2$. Natarajan and Mohankumar employed a variety of quadrature schemes that took care of the singularities which resulted in reduced number of quadrature terms. Trapezoidal and Gauss–Legendre schemes with the correction terms for the poles were employed [14,21,22]. The clustering of the quadrature nodes that is inherent in numerical integration methods like the TANH, the IMT and the DE schemes was also profitably exploited to handle the singularities of these integrands [23,24].

## 2. Existing methods for the FDI and the GFDI evaluation

In this section, we discuss our earlier methods for the evaluation of the FDI and the GFDI. This will help us to identify the improvements that are needed for our earlier algorithms. In addition, it provides the necessary background for the new algorithms that are presented in the next two sections.

For the FDI needed in semiconductor applications, typical $k$ values belong to the set $\{-(1/2), (1/2), (3/2), (5/2)\}$ and typical $\eta$ values lie in the range $[-10, 50]$. For this range of parameters, we first make a change of the integration variable from $t$ to $x$ defined by $t = x^2$ and this has the positive effect of removing the branch point singularity at the origin. The resulting FDI (with an integrand that is even) and its new singularities $\{z_j\}$ are given below and for the sake of simplicity, we omit the factor $\frac{1}{\Gamma(1+k)}$.

$$F_k(\eta) = \int_{-\infty}^{\infty}\frac{x^{2k+1}\,dx}{e^{x^2-\eta}+1} \tag{3}$$

$$z_j = \pm\sqrt{\eta + i(2j+1)\pi}; \quad j = 0, \pm1, \pm2, \ldots. \tag{4}$$

A simple trapezoidal scheme with residue correction for the poles $\{z_j\}$ of this integrand can yield a double precision accuracy (i.e a minimum of 14 digit accuracy) with a maximum of 29 quadrature terms and 7 residue correction terms with $\eta$ in the range $[-10, 50]$ and for half-integer $k$ values. This accuracy stemming from this very modest computational requirements must be sufficient for routine estimation of the FDI. These results were reported in Mohankumar et al. [14]. Table 1 gives sample values of the FDI without its gamma function term. An algorithm based on this scheme is available as a matlab routine called fermi.m that can be freely downloaded by users [25]. For the sake of completion, a brief derivation of this trapezoidal scheme, the residue correction and its discretization error are outlined in Appendix A.

For the above mentioned scheme, it must be noted that the number of trapezoidal terms is proportional to $\eta^{1/2}$. Hence, for large $\eta$ values, this implies more computational cost. To overcome this aspect, the recent quadrature methods, namely, the TANH, the IMT and the Double Exponential (DE) schemes were employed. Essentially, all the three methods are just trapezoidal schemes *after* a specific change of integration variable. We only outline the DE scheme since it is superior to the other two methods. The details of the TANH and the IMT methods can be found in Natarajan et al. [23]. With $t$ and $u$ as the old and the new variables, the DE transformation introduced by Takahasi and Mori [26,27] is defined as follows.

$$t \in (a, b); \quad u \in (-\infty, \infty) \tag{5}$$

$$t_k = \phi(u_k) = (1/2)(b+a) + (1/2)(b-a)\tanh\left(\frac{\pi}{2}\sinh(u_k)\right) \tag{6}$$

$$\frac{dt}{du} = \phi'(u) = \frac{\pi\,(b-a)}{4}\,\mathrm{sech}^2\left[\frac{\pi}{2}\sinh(u)\right]\cosh(u) \tag{7}$$

$$u_k = kh, \quad k = 0, \pm1, \pm2, \ldots. \tag{8}$$

$t_k$, the images of the equi-spaced nodes $u_k$ get clustered at the ends $t = a$ and $t = b$ of the old interval $(a, b)$. This specific

**Table 1**
The FDI values by the trapezoidal scheme with residue correction.

| $k$ | $\eta$ | $n$ | npole | Integral |
|---|---|---|---|---|
| −0.5 | −10 | 20 | 0 | 8.046669716113732e−5 |
| −0.5 | 10 | 24 | 4 | 6.29713724453385 |
| −0.5 | 20 | 25 | 5 | 8.93497266616697 |
| −0.5 | 30 | 26 | 5 | 10.94942130440661 |
| −0.5 | 40 | 27 | 6 | 12.64585068849794 |
| −0.5 | 50 | 29 | 7 | 14.13980529101102 |
| 2.5 | −10 | 20 | 0 | 1.508792951864649e−4 |
| 2.5 | 10 | 24 | 4 | 1.034684254181534e3 |
| 2.5 | 20 | 25 | 5 | 1.059063917661438e4 |
| 2.5 | 30 | 26 | 5 | 4.292925758509994e4 |
| 2.5 | 40 | 27 | 6 | 1.166899210191305e5 |
| 2.5 | 50 | 29 | 7 | 2.539925685770061e5 |

property is exploited to take care of the requirement of increased sampling near the singularities of the integrand in the following way. First due to finite machine precision, we truncate the semi-infinite interval $(0, \infty)$ to $(0, \eta + m)$ where $m$ lies in the range [40, 80]. Then, we split this truncated range into two intervals $I_1 = (0, \eta)$ and $I_2 = (\eta, \eta + m)$ and the interval $(a, b)$ defined in Eq. (5) is either $I_1$ or $I_2$. Since the point $t = \eta$ is an end point for both $I_1$ and $I_2$, there will be a node clustering at $t = \eta$ once we apply the DE transformation to both these intervals. This node clustering results in increased sampling around the point $t = \eta$ in the vicinity of which the singularities $t_j = \eta + i(2j + 1)\pi,\ j = 0, \pm1, \pm2, \ldots$ are located. This results in a drastic reduction of the number of quadrature steps and high accuracy [23,22]. For a convergence of at least 14 digits, we need about 180 quadrature steps in each of the intervals $I_1$ and $I_2$. In the next section, we show a new method to cut down this effort almost by a factor 2 for a near double precision evaluation of the GFDI for routine evaluation.

## 3. A new GFDI evaluation algorithm with near double precision accuracy

The first step is to split the interval $(0, \eta + m)$ as two intervals, $I_1 = (0, \eta - m)$ and $I_2 = (\eta - m, \eta + m)$. With $m$ set to 55, we use a Gauss–Legendre (GL) quadrature over the interval $I_2$ with a residue correction strategy for the poles of the integrand. For this correction, we use a continued fraction algorithm [24]. When the interval $(\eta - m, \eta + m)$ is mapped onto $(-1, 1)$, the poles of the integrand $t_j = \eta + i(2j + 1)\pi,\ j = 0, \pm1, \pm2, \ldots$ are transformed to $\zeta_j = i(2j + 1)\pi/m,\ j = 0, \pm1, \pm2, \ldots$. That is, the transformed poles are purely imaginary. For the interval $I_1$, we employ the DE scheme. For the computation, the parameter $\eta$ ranges from 100 to 50 000 and the parameter $\theta$ can take values 50 or 1. The results are shown in Table 2. The number of DE quadrature steps is 140. A GL scheme of order 21 with 7 residue correction terms is sufficient for the second interval. The maximum absolute relative error is $1.2d - 13$ for $\eta = 50\,000$ and in the remaining cases, the actual errors are much less. When we employ the DE scheme for *both* the intervals $I_1$ and $I_2$ with $I_1 = (0, \eta)$ and $I_2 = (\eta, \eta + m)$ as indicated earlier [22], we require about 160 quadrature steps in each interval and hence a total of about 320 quadrature terms is required. Thus, the method just outlined cuts the computational effort roughly by a factor two and hence for routine evaluation close to double precision accuracy, this new approach that combines the DE and GL schemes is more suitable. It must be kept in mind that the residue correction for the poles of the GFDI plays a key role in the reduction of the computational effort.

## 4. A new algorithm for a very accurate evaluation of the GFDI for $\eta > 1000$

For repeated and very quick evaluation of the FDI and the GFDI, rational approximations are desirable. These rational

**Table 2**
Near double precision values of the GFDI.

| $k$ | $\eta$ | $\theta$ | Integral | Rel err. |
|---|---|---|---|---|
| 0.5 | 50000. | 1. | 0.35364187040847D+05 | 0.2709D−13 |
| −0.5 | 50000. | 50. | 0.250001642494870D+06 | 0.2731D−13 |
| −0.5 | 30000. | 1. | 0.21221690213566D+05 | 0.4996D−14 |
| −0.5 | 30000. | 50. | 0.15000159141232D+06 | 0.4663D−14 |
| −0.5 | 10000. | 1. | 0.70787777771944D+04 | 0.4219D−14 |
| −0.5 | 10000. | 50. | 0.50001481551154D+05 | 0.3109D−14 |
| −0.5 | 5000. | 1. | 0.35427537775051D+04 | 0.8882D−15 |
| −0.5 | 5000. | 50. | 0.25001412236531D+05 | 0.8882D−15 |
| −0.5 | 2000. | 1. | 0.14207856243402D+04 | 0.1110D−15 |
| −0.5 | 2000. | 50. | 0.10001320607723D+05 | 0.8882D−15 |
| −0.5 | 1000. | 1. | 0.71318888985478D+03 | 0.4441D−15 |
| −0.5 | 1000. | 50. | 0.50012512933820D+04 | 0.1110D−14 |
| −0.5 | 500. | 1. | 0.35914571973243D+03 | 0.0000D+00 |
| −0.5 | 500. | 50. | 0.25011819791704D+04 | 0.2220D−15 |
| −0.5 | 200. | 1. | 0.14636680265558D+03 | 0.7772D−15 |
| −0.5 | 200. | 50. | 0.10010903496424D+04 | 0.6661D−15 |
| −0.5 | 100. | 1. | 0.75167663793253D+02 | 0.9992D−15 |
| −0.5 | 100. | 50. | 0.50102102757876D+03 | 0.1443D−14 |
| 0.5 | 50000. | 1. | 0.88391882909172D+09 | 0.5440D−13 |
| 0.5 | 50000. | 50. | 0.62500050082097D+10 | 0.5529D−13 |
| 0.5 | 30000. | 1. | 0.31821926218746D+09 | 0.9326D−14 |
| 0.5 | 30000. | 50. | 0.22500030082103D+10 | 0.9104D−14 |
| 0.5 | 10000. | 1. | 0.35362407965613D+08 | 0.6883D−14 |
| 0.5 | 10000. | 50. | 0.25000100821135D+09 | 0.7772D−14 |
| 0.5 | 5000. | 1. | 0.88423683822408D+07 | 0.1332D−14 |
| 0.5 | 5000. | 50. | 0.62500508212048D+08 | 0.1332D−14 |
| 0.5 | 2000. | 1. | 0.14156261832902D+07 | 0.4441D−15 |
| 0.5 | 2000. | 50. | 0.10000208212964D+08 | 0.5551D−15 |
| 0.5 | 1000. | 1. | 0.35425914961760D+06 | 0.4441D−15 |
| 0.5 | 1000. | 50. | 0.25001082136574D+07 | 0.2220D−15 |
| 0.5 | 500. | 1. | 0.88740797995433D+05 | 0.1110D−15 |
| 0.5 | 500. | 50. | 0.62505821435052D+06 | 0.6661D−15 |
| 0.5 | 200. | 1. | 0.14282776850221D+05 | 0.5551D−15 |
| 0.5 | 200. | 50. | 0.10002821526679D+06 | 0.1887D−14 |
| 0.5 | 100. | 1. | 0.36057078101240D+04 | 0.1221D−14 |
| 0.5 | 100. | 50. | 0.25018215959958D+05 | 0.7772D−15 |
| 1.5 | 50000. | 1. | 0.29463666531558D+14 | 0.8282D−13 |
| 1.5 | 50000. | 50. | 0.20833345915575D+15 | 0.8282D−13 |
| 1.5 | 30000. | 1. | 0.63642792879173D+13 | 0.1354D−13 |
| 1.5 | 30000. | 50. | 0.45000045493450D+14 | 0.1454D−13 |
| 1.5 | 10000. | 1. | 0.23573763546629D+12 | 0.1110D−13 |
| 1.5 | 10000. | 50. | 0.16666718311502D+13 | 0.1021D−13 |
| 1.5 | 5000. | 1. | 0.29471631252003D+11 | 0.2220D−14 |
| 1.5 | 5000. | 50. | 0.20833466557520D+12 | 0.2665D−14 |
| 1.5 | 2000. | 1. | 0.18870362459970D+10 | 0.2220D−15 |
| 1.5 | 2000. | 50. | 0.13333566230179D+11 | 0.8882D−15 |
| 1.5 | 1000. | 1. | 0.23605779007707D+09 | 0.6661D−15 |
| 1.5 | 1000. | 50. | 0.16667331151720D+10 | 0.2220D−15 |
| 1.5 | 500. | 1. | 0.29552160576083D+08 | 0.0000D+00 |
| 1.5 | 500. | 50. | 0.20835405766836D+09 | 0.4441D−15 |
| 1.5 | 200. | 1. | 0.19001577547295D+07 | 0.1221D−14 |
| 1.5 | 200. | 50. | 0.13338623166142D+08 | 0.5551D−15 |
| 1.5 | 100. | 1. | 0.23943781385534D+06 | 0.2220D−15 |
| 1.5 | 100. | 50. | 0.16688116653945D+07 | 0.1554D−14 |
| 2.5 | 50000. | 1. | 0.11048838166683D+19 | 0.1099D−12 |
| 2.5 | 50000. | 50. | 0.78125042283505D+19 | 0.1097D−12 |
| 2.5 | 30000. | 1. | 0.14319549013278D+18 | 0.1910D−13 |
| 2.5 | 30000. | 50. | 0.10125009222062D+19 | 0.1910D−13 |
| 2.5 | 10000. | 1. | 0.1768002986519 1D+16 | 0.1421D−13 |
| 2.5 | 10000. | 50. | 0.12500035800688D+17 | 0.1465D−13 |
| 2.5 | 5000. | 1. | 0.11051498017274D+15 | 0.3997D−14 |
| 2.5 | 5000. | 50. | 0.78125478350614D+15 | 0.2665D−14 |
| 2.5 | 2000. | 1. | 0.28303259988149D+13 | 0.5551D−15 |
| 2.5 | 2000. | 50. | 0.20000365361426D+14 | 0.2220D−15 |
| 2.5 | 1000. | 1. | 0.17701571289681D+12 | 0.3331D−15 |
| 2.5 | 1000. | 50. | 0.12500580072302D+13 | 0.4441D−15 |
| 2.5 | 500. | 1. | 0.11078835747196D+11 | 0.5551D−15 |
| 2.5 | 500. | 50. | 0.78135335265741D+11 | 0.0000D+00 |
| 2.5 | 200. | 1. | 0.28486137798016D+09 | 0.6661D−15 |
| 2.5 | 200. | 50. | 0.20012537297285D+10 | 0.8882D−15 |
| 2.5 | 100. | 1. | 0.17946771869476D+08 | 0.1443D−14 |
| 2.5 | 100. | 50. | 0.12528015816436D+09 | 0.9992D−15 |

**Table 3**
High precision values of the GFDI for $k = -0.5, 0.5$.

| $k$ | $\eta$ | $\theta$ | Integral | Rel err. |
|---|---|---|---|---|
| −0.50 | 50000.0 | 1.0 | 0.353641870408466703558 5Q+05 | 0.237D−22 |
| −0.50 | 50000.0 | 50.0 | 0.250001642494866974036 1Q+06 | 0.668D−23 |
| −0.50 | 30000.0 | 1.0 | 0.212216902135662435012 4Q+05 | 0.182D−24 |
| −0.50 | 30000.0 | 50.0 | 0.150001591412317813790 3Q+06 | 0.144D−24 |
| −0.50 | 10000.0 | 1.0 | 0.707877777719439781401 3Q+04 | 0.109D−26 |
| −0.50 | 10000.0 | 50.0 | 0.500014815511541513965 2Q+05 | 0.326D−24 |
| −0.50 | 5000.0 | 1.0 | 0.354275377750507651419 2Q+04 | 0.119D−25 |
| −0.50 | 5000.0 | 50.0 | 0.250014122365311603211 1Q+05 | 0.653D−24 |
| −0.50 | 2000.0 | 1.0 | 0.142078562434017388543 3Q+04 | 0.960D−25 |
| −0.50 | 2000.0 | 50.0 | 0.100013206077234275059 2Q+05 | 0.163D−23 |
| −0.50 | 1000.0 | 1.0 | 0.713188889854784192422 2Q+03 | 0.776D−24 |
| −0.50 | 1000.0 | 50.0 | 0.500125129338199576967 6Q+04 | 0.324D−23 |
| −0.50 | 500.0 | 1.0 | 0.359145719732429322068 2Q+03 | 0.235D−23 |
| −0.50 | 500.0 | 50.0 | 0.250118197917043553718 1Q+04 | 0.654D−23 |
| −0.50 | 200.0 | 1.0 | 0.146366802655583977311 2Q+03 | 0.987D−21 |
| −0.50 | 200.0 | 50.0 | 0.100109034964236913589 1Q+04 | 0.375D−22 |
| −0.50 | 100.0 | 1.0 | 0.751676637932535166912 2Q+02 | 0.198D−19 |
| −0.50 | 100.0 | 50.0 | 0.501021027578762407340 2Q+03 | 0.430D−21 |
| 0.50 | 50000.0 | 1.0 | 0.883918829091723556912 4Q+09 | 0.103D−23 |
| 0.50 | 50000.0 | 50.0 | 0.625000500820974538537 1Q+10 | 0.476D−25 |
| 0.50 | 30000.0 | 1.0 | 0.318219262187461224251 1Q+09 | 0.236D−23 |
| 0.50 | 30000.0 | 50.0 | 0.225000300821025621073 0Q+10 | 0.870D−24 |
| 0.50 | 10000.0 | 1.0 | 0.353624079656130200526 4Q+08 | 0.281D−22 |
| 0.50 | 10000.0 | 50.0 | 0.250001008211354821699 6Q+09 | 0.941D−23 |
| 0.50 | 5000.0 | 1.0 | 0.884236838224075974477 3Q+07 | 0.133D−21 |
| 0.50 | 5000.0 | 50.0 | 0.625005082120479669295 3Q+08 | 0.977D−22 |
| 0.50 | 2000.0 | 1.0 | 0.141562618329024847629 3Q+07 | 0.343D−21 |
| 0.50 | 2000.0 | 50.0 | 0.100002082129642520068 9Q+08 | 0.186D−21 |
| 0.50 | 1000.0 | 1.0 | 0.354259149617600528969 3Q+06 | 0.398D−20 |
| 0.50 | 1000.0 | 50.0 | 0.250010821365739042128 9Q+07 | 0.100D−20 |
| 0.50 | 500.0 | 1.0 | 0.887407979954325489111 9Q+05 | 0.131D−19 |
| 0.50 | 500.0 | 50.0 | 0.625058214350522537109 8Q+06 | 0.603D−20 |
| 0.50 | 200.0 | 1.0 | 0.142827768502209841759 1Q+05 | 0.483D−19 |
| 0.50 | 200.0 | 50.0 | 0.100028215266787817077 4Q+06 | 0.224D−20 |
| 0.50 | 100.0 | 1.0 | 0.360570781012397979999 6Q+04 | 0.220D−18 |
| 0.50 | 100.0 | 50.0 | 0.250182159599584355707 0Q+05 | 0.198D−18 |

approximation algorithms for the FDI were provided by Cody, Macleod and Antia [11,12,28]. The rational approximation of Antia is about 12 digits accurate [28]. In order to generate this 12 digit accurate approximation, he needs a 20 digit accurate reference integral values that are generated by a repeated Simpson quadrature involving about 2000 function evaluations. In the following, we provide an efficient algorithm based on the DE method and the GL scheme with residue correction where we get a relative error smaller than $1.d(-20)$. The computational cost is moderate since the number of quadrature terms is about 270 or less for $\eta > 1000$. For $\eta < 1000$, a quadruple precision version of our earlier trapezoidal scheme with pole correction will meet the requirement.

As in the double precision case that has been described in Section 3, the key to the high precision evaluation is the use of the GL quadrature with the residue correction over the interval $(\eta - m, \eta + m)$. This is done with a 35 point GL quadrature with 14 residue correction terms and using a quadruple precision Fortran that guarantees a minimum of 32 digit computational accuracy. Computational experiments indicate that the *remaining interval* namely $(0, \eta - m)$ be better split into two intervals. This implies that the total interval $(0, \eta + m)$ is split into three intervals, $I_1 = (0, X)$, $I_2 = (X, \eta - m)$ and $I_3 = (\eta - m, \eta + m)$. For both $I_2$ and $I_3$, we employ the GL integration. For the interval $I_1$, we have the following variation. For $k = -0.5, 0.5$, we employ the GL quadrature with $X = 40$ and for $k = 1.5, 2.5$, we employ the DE quadrature with $X = 5$. It must be emphasized that precise values of the break point $X$ are not critical. Figs. 1–4 gives a plot of the GFD integrand (after the transformation $t = x^2$), given by $[2x^{2k+1}\sqrt{1+\theta x^2/2}]/[e^{x^2-\eta}+1]$ as a function of $x$ for $k = -(1/2), (1/2), (3/2), (5/2)$ with $\theta = 50$ and $\eta = 100$. When $k$ is either $-(1/2)$ or $(1/2)$, we find that the integrand varies rather smoothly away from the point $t = \eta$, suggesting the use of the GL quadrature in $I_1$. When $k$ is either $(3/2)$ or $(5/2)$, the function variation is *relatively* sharp compared to the previous case and hence we use the DE quadrature in $I_1$.

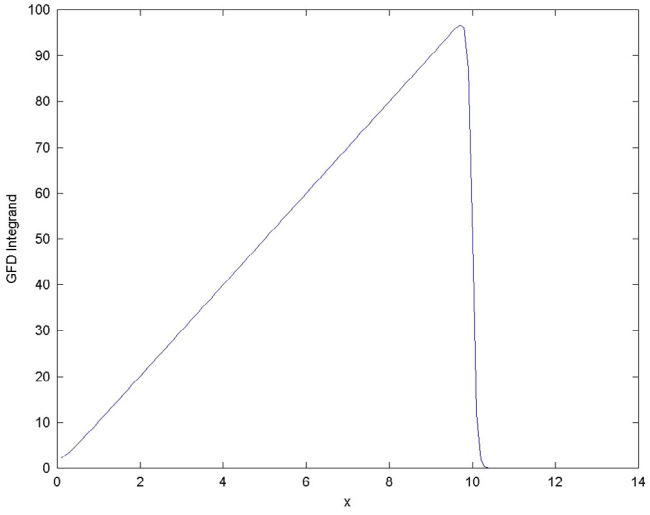


**Fig. 1.** Plot of the GFD integrand with $\theta = 50$ and $k = -0.5$.

The results of these computations are indicated in Tables 3 and 4. The parameter $m$ is set to 80. For $k = -0.5, 0.5$, the GL quadrature order for $I_1$ and $I_2$ are 80 and 60 respectively. Hence, the total number of quadrature terms in this case is 175 (=80+60+35) and in addition we need another 14 residue correction terms. For $k = 1.5, 2.5$, the order of the DE quadrature for $I_1$ is 160 and a GL quadrature of order 76 is employed for $I_2$. In this case, the number

**Table 4**
High precision values of the GFDI for $k = 1.5, 2.5$.

| $k$ | $\eta$ | $\theta$ | Integral | Rel err. |
|---|---|---|---|---|
| 1.50 | 50000.0 | 1.0 | 0.29463666531555763612430Q+14 | 0.182D−23 |
| 1.50 | 50000.0 | 50.0 | 0.20833345915575505315427Q+15 | 0.304D−23 |
| 1.50 | 30000.0 | 1.0 | 0.63642792879172603052 25Q+13 | 0.371D−23 |
| 1.50 | 30000.0 | 50.0 | 0.45000045493450384829 49Q+14 | 0.596D−23 |
| 1.50 | 10000.0 | 1.0 | 0.23573763546629190036 93Q+12 | 0.980D−22 |
| 1.50 | 10000.0 | 50.0 | 0.16666718311502381045 40Q+13 | 0.296D−22 |
| 1.50 | 5000.0 | 1.0 | 0.29471631252002643420 04Q+11 | 0.450D−21 |
| 1.50 | 5000.0 | 50.0 | 0.20833466557520141493 20Q+12 | 0.296D−21 |
| 1.50 | 2000.0 | 1.0 | 0.18870362459969780609 62Q+10 | 0.189D−20 |
| 1.50 | 2000.0 | 50.0 | 0.13333566230179391159 56Q+11 | 0.124D−20 |
| 1.50 | 1000.0 | 1.0 | 0.23605779007706716472 88Q+09 | 0.516D−20 |
| 1.50 | 1000.0 | 50.0 | 0.16667331151720421479 16Q+10 | 0.549D−20 |
| 1.50 | 500.0 | 1.0 | 0.29552160576082616924 26Q+08 | 0.193D−19 |
| 1.50 | 500.0 | 50.0 | 0.20835405766836071091 13Q+09 | 0.311D−19 |
| 1.50 | 200.0 | 1.0 | 0.19001577547295125690 93Q+07 | 0.854D−19 |
| 1.50 | 200.0 | 50.0 | 0.13338623166141841231 16Q+08 | 0.126D−18 |
| 1.50 | 100.0 | 1.0 | 0.23943781385533674087 18Q+06 | 0.573D−18 |
| 1.50 | 100.0 | 50.0 | 0.16688116653944634257 48Q+07 | 0.105D−17 |
| 2.50 | 60000.0 | 1.0 | 0.22910768946583796599 70Q+19 | 0.582D−23 |
| 2.50 | 60000.0 | 50.0 | 0.16200007288824659407 03Q+20 | 0.210D−23 |
| 2.50 | 50000.0 | 1.0 | 0.11048838166683024525 37Q+19 | 0.432D−23 |
| 2.50 | 50000.0 | 50.0 | 0.78125042283504606806 36Q+19 | 0.151D−23 |
| 2.50 | 30000.0 | 1.0 | 0.14319549013277681495 97Q+18 | 0.229D−22 |
| 2.50 | 30000.0 | 50.0 | 0.10125009222061698294 76Q+19 | 0.143D−22 |
| 2.50 | 10000.0 | 1.0 | 0.17680029865191067057 13Q+16 | 0.138D−21 |
| 2.50 | 10000.0 | 50.0 | 0.12500035800687780494 12Q+17 | 0.924D−22 |
| 2.50 | 5000.0 | 1.0 | 0.11051498017273562881 01Q+15 | 0.400D−21 |
| 2.50 | 5000.0 | 50.0 | 0.78125478350614358913 67Q+15 | 0.649D−21 |
| 2.50 | 2000.0 | 1.0 | 0.28303259988148782780 95Q+13 | 0.299D−20 |
| 2.50 | 2000.0 | 50.0 | 0.20000365361425511506 16Q+14 | 0.270D−20 |
| 2.50 | 1000.0 | 1.0 | 0.17701571289680744736 24Q+12 | 0.827D−20 |
| 2.50 | 1000.0 | 50.0 | 0.12500580072301631896 44Q+13 | 0.214D−19 |
| 2.50 | 500.0 | 1.0 | 0.11078835747195889806 71Q+11 | 0.517D−19 |
| 2.50 | 500.0 | 50.0 | 0.78135335265741074119 22Q+11 | 0.137D−19 |
| 2.50 | 200.0 | 1.0 | 0.28486137798015714911 46Q+09 | 0.996D−18 |
| 2.50 | 200.0 | 50.0 | 0.20012537297284586537 36Q+10 | 0.818D−18 |
| 2.50 | 100.0 | 1.0 | 0.17946771869476107109 55Q+08 | 0.156D−16 |
| 2.50 | 100.0 | 50.0 | 0.12528015816436262855 27Q+09 | 0.166D−16 |

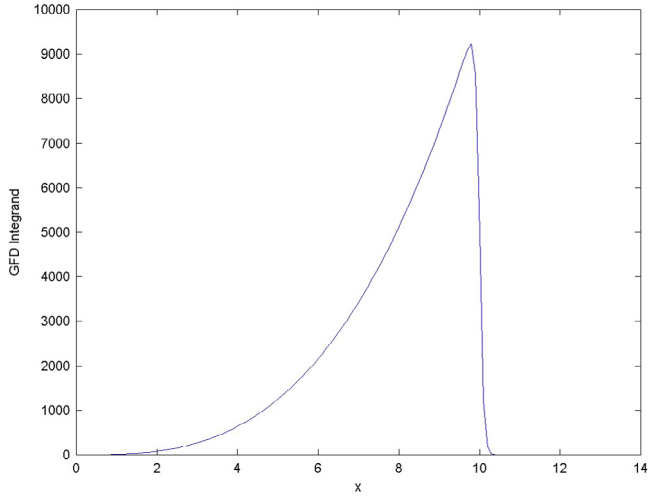


**Fig. 2.** Plot of the GFD integrand with $\theta = 50$ and $k = 0.5$.

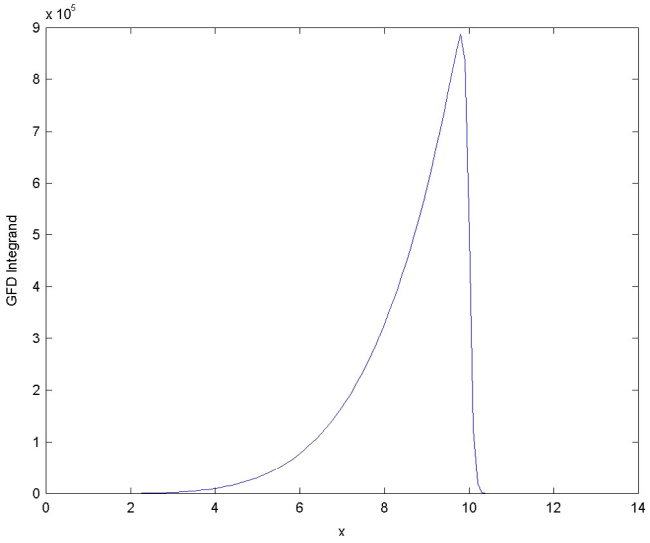


**Fig. 3.** Plot of the GFD integrand with $\theta = 50$ and $k = 1.5$.

of quadrature terms is 271 (=160 + 76 + 35) with the addition of another 14 residue related terms. If we confine to the results in these tables with $\eta > 1000$, then the absolute relative error for the $k$ values considered is found to be of the order of $1.d(-20)$ or less. Thus, the three interval evaluation strategy guarantees a very precise evaluation of the GFDI for $\eta > 1000$. This must be contrasted with the results quoted by Antia where one needs something like a 2000 function evaluations to realize a similar accuracy [28].

The loss of accuracy is more pronounced for small $\eta$ values like $\eta = 100$ and for large $\theta$ values like $\theta = 50$ as we see in Tables 3 and 4. This is an inherent complication that can be easily explained as a result of the influence of the branch point singularity $t_0 = -(\theta/2)$ which stems from the term $\sqrt{1+\theta t/2}$. This singularity affects the quadrature over *all* the three intervals acutely for lower values of $\eta$ and higher values of $\theta$. Unlike the residue correction for the pole $t_j = \eta + i(2j+1)\pi$ of the integrand, we have no

**Table 5**
High precision the GFDI values for larger $\theta$ values with $\eta \leq 1000$ by Trapezoidal scheme with pole correction.

| $k$ | $\eta$ | $\theta$ | Integral | $n$ | $np$ | Rel. err |
|---|---|---|---|---|---|---|
| −0.5 | 100. | 1. | 0.75167663793253511669276Q+02 | 70 | 4 | 0.7374D−21 |
| −0.5 | 100. | 50. | 0.50102102757876240473392Q+03 | 404 | 1 | 0.2400D−20 |
| −0.5 | 200. | 1. | 0.14636680265558397731 14Q+03 | 86 | 6 | 0.6838D−21 |
| −0.5 | 200. | 50. | 0.10010903496423691358 90Q+04 | 494 | 1 | 0.1201D−20 |
| −0.5 | 500. | 1. | 0.35914571973242932206 89Q+03 | 163 | 7 | 0.1762D−20 |
| −0.5 | 500. | 50. | 0.25011819791704355371 67Q+04 | 653 | 2 | 0.5844D−20 |
| −0.5 | 1000. | 1. | 0.71318888985478419242 19Q+03 | 147 | 16 | 0.4711D−21 |
| −0.5 | 1000. | 50. | 0.50012512933819957696 53Q+04 | 872 | 3 | 0.4634D−20 |
| 0.5 | 100. | 1. | 0.36057078101239798007 74Q+04 | 70 | 4 | 0.3969D−20 |
| 0.5 | 100. | 50. | 0.25018215959958435575 69Q+05 | 336 | 1 | 0.1111D−20 |
| 0.5 | 200. | 1. | 0.14282776850220984176 63Q+05 | 69 | 8 | 0.2178D−20 |
| 0.5 | 200. | 50. | 0.10002821526678781707 80Q+06 | 376 | 1 | 0.3927D−20 |
| 0.5 | 500. | 1. | 0.88740797995432548912 02Q+05 | 97 | 12 | 0.3747D−20 |
| 0.5 | 500. | 50. | 0.62505821435052253711 49Q+06 | 510 | 2 | 0.2032D−20 |
| 0.5 | 1000. | 1. | 0.35425914961760052897 23Q+06 | 118 | 20 | 0.4526D−20 |
| 0.5 | 1000. | 50. | 0.25001082136573904213 07Q+07 | 631 | 4 | 0.6345D−20 |
| 1.5 | 100. | 1. | 0.23943781385533674100 86Q+06 | 51 | 6 | 0.1605D−20 |
| 1.5 | 100. | 50. | 0.16688116653944634274 97Q+07 | 257 | 1 | 0.1404D−20 |
| 1.5 | 200. | 1. | 0.19001577547295125692 54Q+07 | 62 | 9 | 0.7401D−21 |
| 1.5 | 200. | 50. | 0.13338623166141841232 82Q+08 | 288 | 2 | 0.1381D−20 |
| 1.5 | 500. | 1. | 0.29552160576082616924 81Q+08 | 89 | 14 | 0.6340D−21 |
| 1.5 | 500. | 50. | 0.20835405766836071091 59Q+09 | 326 | 4 | 0.8845D−20 |
| 1.5 | 1000. | 1. | 0.23605779007706716472 96Q+09 | 110 | 21 | 0.1830D−20 |
| 1.5 | 1000. | 50. | 0.16667331151720421479 20Q+10 | 414 | 6 | 0.3567D−20 |
| 2.5 | 100. | 1. | 0.17946771869476106829 09Q+08 | 47 | 7 | 0.4223D−20 |
| 2.5 | 100. | 50. | 0.12528015816436262647 03Q+09 | 176 | 2 | 0.2417D−20 |
| 2.5 | 200. | 1. | 0.28486137798015714883 19Q+09 | 52 | 11 | 0.3727D−20 |
| 2.5 | 200. | 50. | 0.20012537297284586521 10Q+10 | 173 | 3 | 0.5745D−20 |
| 2.5 | 500. | 1. | 0.11078835747195889806 24Q+11 | 62 | 20 | 0.8806D−20 |
| 2.5 | 500. | 50. | 0.78135335265741074118 45Q+11 | 195 | 6 | 0.3866D−20 |
| 2.5 | 1000. | 1. | 0.17701571289680744736 43Q+12 | 82 | 28 | 0.1997D−20 |
| 2.5 | 1000. | 50. | 0.12500580072301631896 73Q+13 | 221 | 11 | 0.1868D−20 |

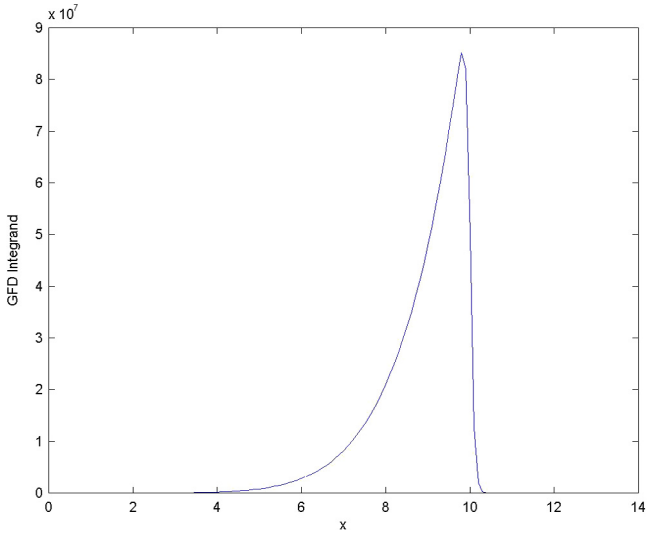


**Fig. 4.** Plot of the GFD integrand with $\theta = 50$ and $k = 2.5$.

way of correcting the quadrature sum for the effect of this branch point. In Appendix B, we deal with this in more detail and we give a criterion for excluding this branch point from our region of analyticity for the GL integration over $(\eta - m, \eta + m)$. Another way of understanding the loss of accuracy for these parameter values is as follows. The third interval $I_3$ combines *both* the tail portion of the integrand where it dies off exponentially and the region in the vicinity of the poles $t_j = \eta + i(2j+1)\pi$, $j = 0, \pm 1, \pm 2, \ldots$. When we set $m = 80$, for $\eta$ values around 100, this region $I_3$ constitutes the bulk of the total interval $(0, \eta+m)$ and a 35 order GL quadrature will be clearly insufficient. To circumvent this problem, one can reduce $m$ to a lesser value and this in turn will require a separate quadrature for the exponentially dying tail portion of the integrand. This implies that the total interval is split into four subintervals. We do not adopt this strategy here.

**Table 6**
Comparison of the DE and the trapezoidal schemes for a minimum relative accuracy better than $1d(-20)$.

| $k$ | $\eta$ | $\theta$ | $n1$ | $n2$ | $n + np$ |
|---|---|---|---|---|---|
| −0.5 | 1000 | 50 | 520 | 400 | 872+3 |
| 0.5 | 1000 | 50 | 520 | 450 | 631+4 |
| 1.5 | 1000 | 50 | 550 | 400 | 414+6 |
| 1.5 | 100 | 50 | 425 | 425 | 257+1 |
| 2.5 | 100 | 50 | 400 | 400 | 176+2 |

## 5. High precision evaluation of the GFDI for $\eta < 1000$

Due to the effect of this branch point singularity $t_0$ that affects the accuracy of the GFDI evaluation, we need a reliable scheme that covers the $\eta$ interval [0, 1000] and for large values of $\theta$. If $\theta$ is much less than unity, the trapezoidal scheme with residue correction is sufficient as we have seen in the results of Table 1. A quadruple precision evaluation of the same scheme will give a guaranteed precision better than $1.d(-20)$. These results are shown in Table 5. Of course for $\eta = 1000$, $k = -0.5$ and $\theta = 50$, we need 872 quadrature terms plus 3 residue correction terms. This is the largest number in this tabulation and for the same $\eta$ value and other $k$ values we need lesser terms. If $\theta$ is of the order of unity, we need just 147 summation terms for $\eta = 1000$. But this computational effort must be compared with the case where one can split the interval as $(0, \eta)$ and $(\eta, \eta + m)$ and do a DE quadrature over these two subintervals. In Table 6, we compare the number of quadrature terms needed for the Trapezoidal scheme with pole correction and the DE quadrature over $I_1 = (0, \eta)$ and $I_2 = (\eta, \eta + m)$. Here, $n1$ and $n2$ denote the quadrature terms needed by the DE scheme in $I_1$ and $I_2$, respectively. $n$ and $np$ denote the number of quadrature steps and the pole correction terms

needed for the trapezoidal scheme, respectively. We find that the trapezoidal scheme is relatively more economical in this range of $\eta$ and $\theta$ values. Before we close, we need to make two observations. For the trapezoidal method with pole correction, we transform the integrand by making the substitution $t = x^2$. The square root term of the integrand of the GFDI given by $\sqrt{1+\theta t/2}$ gets transformed as $\sqrt{1+\theta x^2/2}$. This implies that the branch point of the transformed integrand is given by $\pm i\sqrt{2/\theta}$. That is, the branch points are purely imaginary. If $\theta$ values are large, then $d$, the width of the strip of analyticity can not exceed $\sqrt{2/\theta}$ and since we set $h = ad$, where $a$ is a constant factor, our step size $h$ is limited by this restriction on $d$. This explains why we need an increased number of quadrature steps for large $\theta$ values for the trapezoidal scheme with pole correction. In addition, we must remember that the number of quadrature steps increases proportionately as $\sqrt{\eta+m}$ for this scheme. The second point to be noted concerns the generation of the reference values for fixing the relative errors. For this purpose, we have used a two range DE quadrature over the intervals $I_1 = (0, \eta)$ and $I_2 = (\eta, \eta + m)$ in quadruple precision. The number of quadrature terms are 1200 and 800 for the intervals $I_1$ and $I_2$ respectively. Hence, to generate 32 digit accurate reference values, we have utilized 2000 quadrature terms in all.

## 6. Summary

We summarize the results here.

(1) For the FDI needed in semiconductor modeling where $\theta$ is zero and $\eta$ lies in the range $[-10, 50]$, the simple trapezoidal integration with pole correction that was indicated earlier [14] can yield double precision accuracy with about 29 quadrature terms and 7 pole correction terms which is adequate for routine use.

(2) For the routine double precision evaluation of the GFDI, the DE quadrature algorithm that splits the range as $I_1 = (0, \eta)$ and $I_2 = (\eta, \eta + m)$ was indicated earlier [22] and this method needs about 320 quadrature steps. This can be replaced by the present algorithm indicated in Section 3 that splits the interval as $(0, \eta - m)$ and $(\eta - m, \eta + m)$. The DE quadrature is employed in the first interval and the GL integration with pole correction is employed in the second interval. This method needs just 161 quadrature terms totally and 7 pole correction terms and offers a minimum of about 13 digits of relative accuracy at **half** the computational cost of the previous algorithm.

(3) For reference purposes and for $\eta > 1000$ and for large $\theta$ values, the algorithm indicated in Section 4 that offers a relative accuracy better than $1.d(-20)$ can be used. The range is split as $I_1 = (0, X)$, $I_2 = (X, \eta - m)$ and $I_3 = (\eta - m, \eta + m)$. The GL quadrature is employed in the last two intervals and in addition, the residue correction is implemented in the third interval. For $k = -0.5, 0.5$, the GL quadrature is implemented in $I_1$. In all we need 175 quadrature terms and 14 residue correction terms. For the case, $k = 1.5, 2.5$, the DE scheme is used in $I_1$. For this we need 271 quadrature terms and 14 residue correction terms. To achieve a similar accuracy by employing the DE scheme after splitting the range as $I_1 = (0, \eta)$ and $I_2 = (\eta, \eta + m)$, we need about 1000 function evaluations. Thus, the present method cuts the computational effort by a factor 1000/285 or 1000/189 which is a significant reduction.

(4) For $\eta < 1000$ and for large $\theta$ values, the inherent nature of the singularities of the integrand of the GFDI warrants more computational effort. For this range, the quadruple precision version of the trapezoidal scheme with pole correction is preferable and this offers an accuracy better than $1.d(-20)$.

Finally, few comments are in order regarding certain analytical evaluation of these integrals found in the literature.

I. The FDI can be evaluated in closed form in a series that involves any one of the following special functions like the Polylogarithm function [29], the Zeta function [30], the incomplete Gamma function and the Error function [15,31]. It must be emphasized that the evaluation of these functions needs special algorithms and this involves an additional computational cost which is not negligible. Also, the rapidity of convergence of these series is another aspect which may not be favorable. On the otherhand, the quadrature based methods presented here involve only very elementary functions and hence there is no extra computational cost and the convergence is robust and rapid.

II. Unlike the FDI, as of now the GFDI does not have a closed form series (except for *very small values of* $\theta$) primarily due to the term $\sqrt{1+\theta t/2}$ of the integrand.

III. The final point concerns the need for *reference numerical values* of high precision. Quite often the FDI integrals are needed in iterative loops in semiconductor device modeling [32]. Similarly, the GFDI is needed in stellar evolution calculations [33]. Here, it is desirable to have good rational approximations for these integrals and this in turn needs very accurate reference values. The minimum accuracy of one part in $10^{20}$ guaranteed by the present algorithms serves these reference purposes quite well.

## 7. Conclusions

We have indicated two new algorithms for the evaluation of the GFDI. The first one provides the values of the GFDI with near double precision accuracy. Compared to its earlier variant, the computational effort is halved and this recipe is suitable for routine evaluation. The second algorithm provides the values of the GFDI with a relative error less than $10^{-20}$. As far as we know, this is the most accurate algorithm at present that can be used for reference purposes. Compared to the methods indicated in the literature, the present one cuts down the function evaluation roughly by a factor 10 which is a significant reduction. These methods are backed by rigorous error estimates [15,26,27,34]. In particular, the effect of the branch point of the integrand of the GFDI is dealt with thoroughly. These methods can be easily extended to the generalized Boltzmann integrals [13]. The Matlab and the Fortran routines that evaluate these integrals can be obtained from the first author upon request.

## Acknowledgments

NMK gratefully acknowledges all the help and encouragement provided by Prof. Tucker Carrington and Prof. Scott M. Auerbach.

## Appendix A. The discretization error of the quadrature

Let $L$ be a contour that encloses the interval $(X_1, X_2)$ as well as an enumerable set of simple poles $z_1, z_2, \ldots$ of the integrand $f(z)$ in the complex $z$ plane. The contour is traversed counter clockwise. We assume that the function $f(z)$ is analytic everywhere within $L$ except at the simple poles $\{z_j\}$. Consider the contour integral $I_0$ that is given below with $\phi_n(x)$ as the basis function.

$$I_0 = \frac{1}{2\pi i}\oint_L \frac{f(z)dz}{(z-x)\phi_n(z)}$$

The singularities of the above integrand consist of the poles $\{z_j\}$ of $f(z)$, $z = x$ and the zeros $\{x_1, x_2, \ldots, x_n\}$ of $\phi_n(x)$. The application of the Residue theorem yields the following expression.

$$I_0 = \frac{f(x)}{\phi_n(x)} + \sum_{i=1}^{n} \frac{f(x_i)}{(x_i - x)\phi_n'(x_i)} + \sum_j \frac{R(z_j)}{(z_j - x)\phi_n(z_j)}$$

$$R(z_j) = \mathrm{Res}[f(z)]_{z=z_j}$$

A slight rearrangement of the above equation leads to the following equation that can be recognized as the expansion of the function $f(x)$ in terms of the basis $\phi_n(x)$.

$$f(x) = \sum_{i=1}^{n} \frac{f(x_i)\,\phi_n(x)}{\phi_n'(x_i)(x-x_i)} + \sum_j \frac{R(z_j)\phi_n(x)}{\phi_n(z_j)(x-z_j)} + I_0\,\phi_n(x)$$

We multiply both sides by the weight function $W(x)$ and then integrate over $(X_1, X_2)$ that leads to the following expression.

$$\int_{X_1}^{X_2} W(x) f(x) dx - \left\{ \sum_1^n w_i f(x_i) + \sum_j \frac{R(z_j)\Phi_n(z_j)}{\phi_n(z_j)} \right\}$$
$$= \frac{(-1)}{2\pi i} \oint_L \frac{f(z)\Phi_n(z)dz}{\phi_n(z)} = E_d \qquad \text{(A.1)}$$

The term $\Phi_n(z)$ is given by

$$\Phi_n(z) = \int_{X_1}^{X_2} \frac{W(x)\,\phi_n(x)dx}{(x-z)}.$$

The first term inside the curly brackets of Eq. (A.1) is the quadrature sum and the second summation term is the correction that is needed for the quadrature sum due to the poles $\{z_j\}$ of $f(z)$. We assume that $f(z)$ is such that the contour integral term $E_d$ that can be identified as the discretization error tends to zero in the limit for large $n$.

**Trapezoidal Integration**

For the trapezoidal integration over $(-\infty, \infty)$, with a step size $h$, the required basis is the sinc function defined by $\phi_k(x) = \frac{\sin(\theta)}{\theta}$, $\theta = [\frac{\pi}{h}(x-kh)]$. Here, the contour $L$ is a rectangle of semi-width $d$ with the real axis symmetrically lying midway between the top and bottom sides of the rectangle [14,22]. With $W(x) = 1$, the function $\Phi$ is simply evaluated in closed form due to the following elementary result [35].

$$\int_{-\infty}^{\infty} \frac{\sin(\alpha x)\,dx}{(x-\beta)} = \pi e^{i\alpha\beta}; \quad \alpha > 0;\ \mathrm{Im}(\beta) > 0.$$

We only quote the final result and the details can be found in Mohankumar et al. [14,22].

$$E_d = \int_{\infty}^{\infty} f(x)\,dx - \left\{ \sum_{k=-\infty}^{\infty} hf(kh) + \sum_j \frac{R(z_j)e^{i\frac{\pi}{h}z_j S(z_j)}}{\sin(\frac{\pi}{h}z_j)} \right\}$$

$$E_d = e^{-2\pi d/h} M; \quad M = \int_{-\infty}^{\infty} [f_1 + f_2]\,dt$$

$$f_1 = \frac{f(t+id)}{e^{-i2\pi t/h} - e^{-2\pi d/h}}; \qquad f_2 = \frac{f(t-id)}{e^{i2\pi t/h} - e^{-2\pi d/h}}$$

Here, $S(z)$ denotes the sign of $\mathrm{Im}(z)$ and the summation over $j$ is the summation over the simple poles of $f(z)$ that lie within the strip of analyticity of semi-width $d$. Also the discretization error $E_d$ is dominated over by the term $e^{-2\pi d/h}$. We take this dominant quantity $e^{-2\pi d/h}$ as an order of magnitude of the error since the other quantity $M$ is bounded as a consequence of the nature of the integrand. The calculation procedure is as follows. We truncate the upper limit of integration to $x_{\max}$. Then, we choose the *ratio* $d/h$ such that the quantity $e^{-2\pi d/h}$ is of the order of the machine precision limit $\epsilon$ (e.g $10^{-16}$ in the double precision case). Having fixed the quantity $d/h$, we then fix $d$ by including a finite number of poles and from this set of poles, the distance of the pole farthest from the real axis is set as $d$. Knowing the ratio $d/h$ and $d$, we get $h$. Since we know both $h$ and $x_{\max}$, the number of quadrature steps $n$ is also fixed finally.

**Gauss–Legendre integration**

For the GL integration, the contour integral error estimate provided by Eq. (A.1) is valid. The interval of interest is $(X_1, X_2) = (-1, 1)$. The integrand function $f(z)$ is analytic but for some simple poles within the region enclosed by $L$. The only difference is that this region is an ellipse that will be defined shortly below.

The following integral is evaluated by the Gauss–Legendre integration with correction for its poles $\{t_j\}$. Here, $m$ values are about 50 and 80 in the double and quadruple precision cases, respectively.

$$I = \int_{\eta-m}^{\eta+m} \frac{t^k \sqrt{1+\theta t/2}\,dt}{e^{t-\eta}+1}$$
$$t_j = \eta + i(2j+1)\pi; \quad j = 0, \pm1, \pm2, \ldots.$$

By the change of variable $t = \eta + mx$, $t \in (\eta - m, \eta + m)$, $x \in (-1, 1)$, this $t$ interval of width $2m$ is mapped on to the Gauss–Legendre interval $(-1, 1)$. The images of the poles of the integrand $t_j = \eta + i(2j+1)\pi$ are purely imaginary and they are given by $\zeta_j = i(2j+1)\pi/m$, $j = 0, \pm1, \pm2, \ldots$. With $W(x) = 1$, the function $\Phi_n(z)$ is given by

$$\Phi_n(z) = \int_{X_1}^{X_2} \frac{W(x)\,\phi_n(x)dx}{(x-z)}$$
$$= \int_1^1 \frac{P_n(x)\,dx}{(x-z)} = -2Q_n(z).$$

Here, $Q_n(z)$ is the second kind Legendre function defined by [36]

$$Q_n(z) = (1/2)\int_1^1 \frac{P_n(x)\,dx}{(z-x)}$$

From Eq. (A.1), it is seen that for the correction to the quadrature sum, we need only the residue term $R(\zeta_j)$ and the quantity $\frac{\Phi_n(\zeta_j)}{\phi_n(\zeta_j)}$. That is, we need only the ratio involving $\Phi_n(\zeta_j)$ and $\phi_n(\zeta_j)$ and not the individual quantities. For the Gauss–Legendre case, this ratio $Q_n(\zeta_j)/P_n(\zeta_j)$ is evaluated accurately by a simple continued fraction algorithm [37]. The details can be found in Mohankumar [38].

Next, we consider the elliptic region enclosed by the contour $L$. This ellipse is defined below [36].

$$z = x + iy \in L; \qquad z = (1/2)(\xi + \xi^{-1});$$
$$\xi = \rho\, e^{i\theta}; \quad \theta \in [0, 2\pi].$$

The ellipse has foci at $z = \pm1$. The semi-axes $a$ and $b$ are given by

$$a = (1/2)(\rho + \rho^{-1}); \qquad b = (1/2)(\rho - \rho^{-1})$$

Let $\rho$ be greater than unity and let $f(z)$ be analytic inside this ellipse but for the poles $\{\zeta_j\}$ which lie on the imaginary axis. Then, the *dominant error* of the Gauss–Legendre quadrature of order $n$ is given by $\pi\rho^{-2n}$ (theorem 3, [34]). In order to apply this theorem, we need to correct the quadrature sum by incorporating the residues of the poles that lie within the ellipse. Let $np$ denote the index of the farthest pole for which residue correction is taken care of. Then, the index $(np+1)$ denotes the first pole starting from which residue evaluation is omitted. Then, we take $b$, the semi-minor axis of the ellipse, as $b = (1/2)|(\zeta_{np} + \zeta_{np+1})|$. In other words

$$b = (1/2)(\pi/m)\,\{[2np-1] + [2(np+1)-1]\}. \qquad \text{(A.2)}$$

Knowing $b$, we can get $\rho$ by solving the equation $b = (1/2)(\rho - \rho^{-1})$. The dominant error of the quadrature is found as $\pi\rho^{-2n}$.

## Appendix B. The effect of the branch point on the quadrature over the interval $I_3$

The results from Tables 2–4 indicate that the accuracy of the quadrature gets impaired for lower values of $\eta$ and higher values of $\theta$, as seen in the case, $\theta = 50$ and $\eta < 1000$. This loss of accuracy can be explained as a consequence of the branch point stemming from the term $\sqrt{1+\theta t/2}$ of the integrand of the GFDI.

In the following, we indicate in detail how this branch point can violate the analyticity assumptions for low values of $\eta$ and higher values of $\theta$. We indicate a criterion for avoiding this violation. For the quadruple precision evaluation, we split the total interval $(0, \eta+m)$ into three intervals and we *do not* invoke the analyticity assumption in the first two intervals. However, for the integration in the third interval $(\eta - m, \eta + m)$, we must ensure that this branch point lies outside the elliptical region we have considered just now. The only singularities that should lie within the ellipse are the finite number of poles from the set given by $t_j = \eta + (2j + 1)\pi;\ j = 0, \pm1, \pm2, \ldots$ and for this finite number of poles, the residue correction is applied to the quadrature sum. This ellipse has foci at $\pm1$. For this ellipse, we fix the semi-minor axis $b$ in terms of the farthest pole for which the residue correction is taken care of. Once we know $b$, the parameter $\rho$ and $a$, the semi-major axis of the ellipse get fixed as follows [34].

$$\rho = b + \sqrt{b^2 + 1}; \qquad a = (1/2)\left(\rho + \frac{1}{\rho}\right) \tag{B.1}$$

The branch point $t_0 = -(2/\theta)$ stems from the square root term $\sqrt{1 + \theta t/2}$. Here $t$ lies in the interval $I_3 = (\eta - m, \eta + m)$ that gets mapped on the Gauss–Legendre interval $(-1, 1)$ by the relation.

$$t = \eta + mx; \quad t \in (\eta - m, \eta + m);\ x \in (-1, 1) \tag{B.2}$$

The image $x_0$ of this branch point $t_0$ under the above mapping is given by $x_0 = -(1/m)[(2/\theta) + \eta]$. For small values of $\eta$ and large values of $\theta$, because of the factor $m$, this branch point that lies on the real axis can enter the interior of the ellipse of analyticity. If it happens, the contour integral error estimate given by Eq. (A.1) is not valid. The poles of the integrand $t_j$ are given by $t_j = \eta + (2j + 1)\pi;\ j = 0, \pm1, \pm2, \ldots$. Under the mapping given by Eq. (B.2), $\zeta_j$, the images $t_j$ are given by $\zeta_j = i(2j + 1)\pi/m,\ j = 0, \pm1, \pm2 \ldots$. These the poles $\zeta_j$ are purely imaginary. As seen earlier, $np$ be the index of the farthest pole on the imaginary axis for which the residue correction is included in the quadrature sum. Then, the index $(np + 1)$ denotes the first pole starting from which residue evaluation is omitted. From Eq. (A.2), we have

$$b = (1/2)(\pi/m)\left\{[2np - 1] + [2(np + 1) - 1]\right\}. \tag{A.2}$$

Once $b$ is known, we can find $\rho$ and hence the semi-major axis can be found from the relation $a = (1/2)(\rho + \rho^{-1})$. Then, if we require the branch point $x_0$ to be located outside the ellipse, we must satisfy the following condition.

$$|x_0| > a.$$

## References


[1] J.S. Blakemore, Solid-State Electron. 25 (1982) 1067.
[2] M. Goano, ACM Trans. Math. Software 21 (1995) 221.
[3] S.A. Wong, S.P. McAlister, Zhan-Ming Li, Solid-State Electron. 37 (1994) 61.
[4] R.B. Dingle, Appl. Sci. Res. B 6 (1957) 225.
[5] R.B. Dingle, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 244 (1958) 456.
[6] J. McDougall, E.C. Stoner, Philos. Trans. Roy. Soc. London Ser. 237 (1939) 67.
[7] M. Goano, Solid-State Electron. 36 (1993) 217.
[8] S.D. Jog, Phys. Lett. A 72 (1979) 303.
[9] D. Bednarczyk, J. Bednarczyk, Phys. Lett. A 64 (1978) 409.
[10] K. Unger, Phys. Status Solidi b 149 (1988) K141.
[11] W.J. Cody, H.C. Thacher Jr., Math. Comp. 21 (1967) 30.
[12] A.J. Macleod, ACM Trans. Math. Software 24 (1998) 1.
[13] V. Bhagat, R. Bhattacharya, D. Roy, Comput. Phys. Comm. 155 (2003) 7.
[14] N. Mohankumar, A. Natarajan, Phys. Status Solidi b 188 (1995) 635.
[15] N. Mohankumar, Comput. Phys. Commun. 176 (2007) 665.
[16] M.R. Spiegel, Theory and Problems of Complex Variables, in: Schaum Outline Series, McGraw-Hill, New York, 1974.
[17] B. Pichon, Comput. Phys. Commun. 55 (1989) 127.
[18] R.P. Sagar, Comput. Phys. Commun. 66 (1991) 271.
[19] W. Gautschi, Comput. Phys. Commun. 74 (1993) 233.
[20] Z. Gong, L. Zejda, W. Dappen, J.M. Aparicio, Comput. Phys. Commun. 136 (2001) 294.
[21] A. Natarajan, N.Mohankumar, Comput. Phys. Commun. 137 (2001) 361.
[22] N. Mohankumar, A. Natarajan, Astrophys. J. 458 (1996) 233.
[23] A. Natarajan, N. Mohankumar, Comput. Phys. Commun. 76 (1993) 48.
[24] N. Mohankumar, T. Kannan, S. Kanmani, Comput. Phys. Commun. 168 (2005) 71.
[25] fermi.m, http://in.mathworks.com/matlabcentral/fileexchange/13616-fermi.
[26] H. Takahasi, M. Mori, Publ. RIMS, Kyoto Univ. 9 (1974) 721.
[27] M. Mori, Proceedings of the International Congress of Mathematicians, Kyoto 1990, Springer-Verlag, Tokyo, 1991, p. 1585.
[28] H.M. Antia, Astrophys. J. Suppl. Ser. 84 (1993) 101.
[29] M, D. Ulrich, W.F. Seng, P.A. Barnes, J. Comp. Electr. 1 (2002) 431.
[30] R. Hill, Amer. J. Phys. 38 (1970) 1440.
[31] J.A. Selvaggi, J.P. Selvaggi, Open Math. J. 5 (2012) 1.
[32] Fermi Level.m, http://in.mathworks.com/matlabcentral/fileexchange/25088-fermilevel.
[33] P.P. Eggleton, J. Faulkner, B.P. Flannery, Astron. Astrophys. 23 (1973) 325.
[34] M.M. Chawla, BIT 10 (1970) 145.
[35] I.S. Gradshteyn, I.W. Ryzhik, Table of Integrals, Series and Products, Academic Press, New York, 1980.
[36] W.W. Bell, Special Functions For Scientists and Engineers, D. Van Nostrand, London, 1968.
[37] A. Erdelyi, W. Magnus, F. Oberhettinger, F.G. Tricomi, Higher Transcendental Functions, Vol. 2, McGraw-Hill, New York, 1953.
[38] N. Mohankumar, Comput. Sci. Eng. 13 (2011) 58.