

A TCAD-Friendly Numerical Treatment of Bandtails in Cryogenic and Amorphous Oxide Semiconductors

M. Thesberg, J. M. Gonzalez-Medina, Z. Stanojevic, M. Kampl, F. Schanovsky, O. Baumgartner, M. Karner
Global TCAD Solutions GmbH
 Vienna, Austria
 Email: m.thesberg@globaltcad.com

Abstract—Bandtails, being spectrums of bound carrier states with energies exponentially-decaying into the bandgap, play a crucial role in transport both in cryogenic silicon (i.e. quantum computing and cryo-electronics), and in increasingly technologically important amorphous semiconductor oxides such as In-Ga-Zn-O (IGZO). Thus, a numerically accurate treatment of bandtails is a must. It is known that the relevant math can be expressed exactly through the use of “hypergeometric functions” and previous work has focused on how these functions can be approximated for compact model implementation. However, no work has considered the TCAD perspective on the relevant calculations where direct computation of hypergeometric functions, even using pre-existing numerical libraries such as BOOST, is both inaccurate and inefficient for the parameter regime relevant to bandtails. Here we present a TCAD-friendly numerical scheme for the evaluation of bandtails that exploits a novel reformulation of the relevant math into a series of parameter regimes where for each a customized optimal scheme has been developed. This approach avoids the direct use of the slow and inaccurate hypergeometric functions entirely and also has an immediate benefit of yielding the values of derivatives without additional effort.

Index Terms—bandtails, quantum computing, cryogenic electronics, IGZO, TCAD, numeric computation, hypergeometric functions

I. INTRODUCTION

Bandtails – which naturally arise in semiconductor systems in the presence of disorder, imperfections and impurity – are especially important in cryogenic systems, like those found in quantum computing, and in amorphous materials like In-Ga-Zn-O (IGZO). The contribution to the carrier density by bandtail carriers is given by the expression obtained by multiplying the exponentially decaying density-of-states (DOS) and the Fermi-Dirac occupancy function:

$$n_b = N_b k_B T_0 \int_{-\infty}^{E_b} \frac{\exp\left(\frac{E-E_0}{k_B T_0}\right)}{1 + \exp\left(\frac{E-E_F}{k_B T}\right)} dE \quad (1)$$

where N_b is a DOS constant, E_0 is the bandtail energy cutoff (for the case of multiple band-tails where not all band tails

This work is supported by the Chips JU project ARCTIC (Project #101139908). That project is supported by the Chips Joint Undertaking and its members (including top-up funding by Belgium, Austria (FFG), Germany, Estonia, Finland, France, Ireland, The Netherlands and Sweden). ARCTIC gratefully acknowledges the support of the Canadian and the Swiss federal governments.

start at the band-edge E_b), k_B is the Boltzmann constant, T the temperature and T_0 is the temperature scale of the bandtail. The key difference between the case of silicon ($T_0 = 50$ K) and IGZO ($T_0 = 770$ K) is the value of T/T_0 at room temperature.

The integral present in (1) (neglecting the prefactors for simplicity) has the closed-form analytical solution [1], [2]:

$${}_2F_1(1, b, 1+b; -|z|), \quad z = \exp\left(\frac{E_b - E_F}{k_B T}\right), \quad b = \frac{T}{T_0} \quad (2)$$

where ${}_2F_1$ is the well-established hypergeometric function. Prima facie this expression would seem to be the end of the story and, though hypergeometric functions may not be appropriate for compact model applications [3], for the case of TCAD it would seem straightforward to use, for example, numerical libraries such as BOOST [4] in C/C++ which have ready-made functions like `hypergeometric_pFq` available.

Typically these libraries ultimately use a power-series summation in small- $|z|$ where for the case of bandtails $|z|$ is usually extremely large (since usually $E_b > E_F$). Furthermore, most known change-of-variable transformations of hypergeometric functions that, for example, reformulate $|z|$ to $1/|z|$ and would thus allow evaluation of large- $|z|$, actually fail for the exact case where the three hypergeometric function parameters (a, b, c) of ${}_2F_1(a, b, c; z)$ meet the condition that $c - (b + a) \in \mathbb{Z}$ (i.e. an integer), which is always true for bandtails. In fact, the only transformation that is possible is DLMF 15.8.1 which takes $|z|$ to $-|z|/(-|z| - 1)$ which maps a large $|z|$ to ~ 1 where the power-series done will have extremely slow convergence (NOTE: all known, pre-established fundamental relations of hypergeometric functions will in this work be quoted along with a reference to the *Digital Library of Mathematical Functions* (DLMF, [5])):

$${}_2F_1(1, b; b+1; z) = (1-z)^{-1} {}_2F_1(1, 1; b+1; \frac{z}{z-1}) \quad (\text{DLMF 15.8.1})$$

Thus, despite the existence of seemingly helpful pre-existing numerical implementations of the hypergeometric function, for the real mathematical case of bandtails their numerical behavior will be pathologically poor.

In this work an alternative numerical scheme is presented that instead of using one general algorithm rather - depending

on the size of $|z|$ - uses different specialized numerical approaches for different regimes of $|z|$, none of which invoke the direct evaluation of a hypergeometric function. The structure of the remaining text is as follows: first a pseudo-algorithm of the novel approach will be given, then its accuracy and performance will be compared to the aforementioned “naive” approach and then finally, in the interest of completeness and reproducibility, the mathematical derivation of the key relations the algorithm exploits will be given in the form of an appendix.

II. METHODS AND RESULTS

As was previously discussed, in order to directly use a pre-existing hypergeometric function from a numerical library, one must use the following algorithm which here we call the “Naive Method”.

Algorithm 1: Evaluate ${}_2F_1(1, b; b+1; -|z|)$ (Naive or BOOST Method)

//x is very, very close to 1 for large |z|
 $x \leftarrow \frac{-|z|}{-|z|-1};$
return $\text{hyp2f1}(1, 1, b+1; x) / (1-z);$

The “Naive Method” will then attempt a small- $|z|$ series expansion for a $|z| \sim 1$. Instead we take the following approach: The most common case is a large $|z|$ (i.e. E_F far below E_b). For this regime we use a newly-derived asymptotic relation (derived in III-B) accurate for large- $|z|$ and small- b . However, as $b = T/T_0$ it is not generally true that b should be small, thus we have also derived a recurrence relation (derived in III-C) that allows a large- b to be recursively reduced to a small- b , thus making the relation valid for all temperatures. For the case of moderately sized $|z|$ we present a well-formulated quadrature approach (derived in III-A). Finally for the case of small- $|z|$ either the quadrature approach can also be used or if a hypergeometric function is available it may also be optimal to use the “Naive Method” in this regime only. A pseudo-code version of the algorithm is given in Algorithm 2 To demonstrate the validity of this algorithm TCAD simulations were performed on a 28 nm FD SOI device using the Minimos-NT [6] within the GTS Framework using default simulation parameters and quasi-Fermi transport (QFT, [7]), an alternative transport formalism to drift-diffusion that performs better in cryogenic or ultra-low-carrier-concentration simulations. However, the specific details of the device and simulation are of minimal importance here as they are kept the same in all simulation with the only point of variation being which algorithm is used to evaluate the bandtail math. Figs. 1, II compare the accuracy of the “New” and “Naive” or “BOOST”-based methods for both the cryo-silicon-like case ($T_0 = 50$ K) and IGZO-like case ($T_0 = 770$ K) using both double (Fig. 1) and long double, a.k.a. long, precision (Fig. II). Cases where no symbols exist indicate that the simulation did not converge. This demonstrates that where the “BOOST” routinely fails to converge at lower temperatures and Fermi

Algorithm 2: Evaluate ${}_2F_1(1, b; b+1; -|z|)$ (New Method)

$x \leftarrow -|z|;$
//If $b \sim 1$ (ϵ is machine epsilon)
if $|b-1| > \epsilon$ **then**
 //Invoke DLMF 15.4.1 directly
 return $\ln(x)/x$
end
//Adjustable cut-off sets when x
//is “too big” for quadrature
 $x_{max} \leftarrow 10^{10};$
if $x < x_{max}$ **then**
 return $\text{quadrature}\left(\int_0^1 \frac{1}{1+xy^{1/b}} dy\right)$
end
//r: accumulates prefactors
//during recursion
 $r \leftarrow b;$
//n: integer such that $b_{new} = b - n$ is
//between 0 and 1
 $n \leftarrow 0;$
while $r > 1$ **do**
 $r \leftarrow r - 1;$
 $n \leftarrow n + 1;$
end
//F: Hold the result but needs to
//be recursed via A.14 before returning
 $F \leftarrow 0;$
//Check for DLMF15.4.1 case again
if $|r-1| > \epsilon$ **then**
 $F \leftarrow \ln(x)/x$
else
 //Use asymptotic relation A.11
 $F = \frac{b\pi}{\sin(\pi b)} x^{-b} - \frac{b}{b-1} x^{-1} + \frac{b}{b-2} x^{-2} - \frac{b}{b-3} x^{-3}$
end
for $i \leftarrow 1$ **to** n **do**
 $b' \leftarrow r + i;$
 $F \leftarrow \frac{b'}{x(b'+1)} (1 - F);$
end
return $F;$

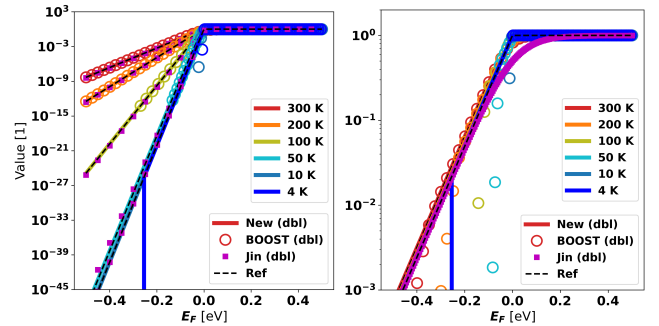


Fig. 1. Integral values for Si-like ($T_0 = 50$ K, left) and IGZO-like ($T_0 = 770$ K, right) vs. E_F for different temperatures and evaluation approaches (double precision)

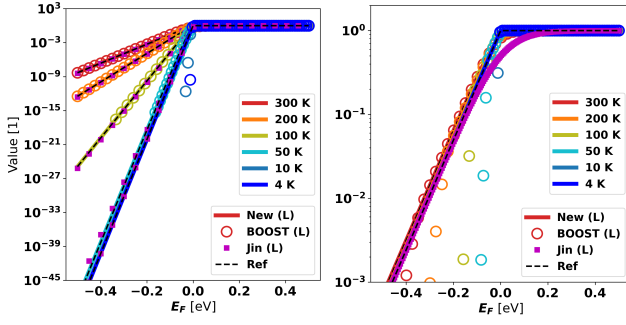


Fig. 2. Same as Fig. 1 but with long double precision

levels, the “Novel” approach does not and is only limited by float precision (i.e. the discontinuity at cryo-temperatures in Fig. 1 is due to numerical underflow and thus doesn’t exist in Fig. II with its increased float precision). Fig. 3 shows the root-mean-square (RMS) error of these results relative to the reference values. Additionally, an alternative simplified approximate analytical approach suitable for compact models called the Jin model [8] is shown for a comparison of accuracy and speed. Finally, a timing performance comparison is also shown in Table I demonstrating that the new approach here has much higher accuracy and numerical efficiency than directly using numerical libraries. An important additional advantage

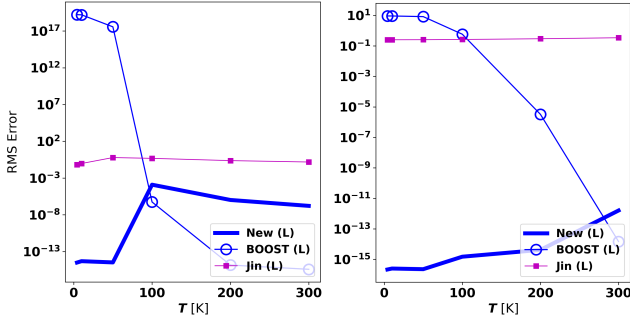


Fig. 3. Root-mean-square (RMS) error of different evaluation approaches vs. temperature for Si-like (left) and IGZO-like (right) bandtails

of the “New” algorithm worth stressing is that the derivatives of the carrier concentration (i.e. $\partial n/\partial E_F$ or $\partial n/\partial V$), which are also essential ingredients of a convergent TCAD simulation, can be obtained without the need to evaluate a second hypergeometric function (see III-D) and instead to reuse the value computed for the carrier concentration. This can effectively cut the numerical burden in half.

TABLE I
SPEED UP OVER BOOST APPROACH

Runtime Enhancement	New Method		Jin Method	
	double	long	double	long
Cryo-Si-like: (Carrier Density + Derivatives)	3,200×	150×	$4 \cdot 10^5 \times$	$3 \cdot 10^5 \times$
IGZO-like:	600×	30×	$2 \cdot 10^5 \times$	$1 \cdot 10^5 \times$

III. MATHEMATICAL APPENDIX

A. Proof of Quadrature Relation

Starting from the relation DLMF 15.6.1 and using that $\Gamma(1) = 1$:

$${}_2F_1(1, b; b+1; -x) = \frac{1}{\Gamma(b)\Gamma((b+1)-b)} \int_0^1 \frac{t^{b-1}(1-t)^{(b+1)-b-1}}{(1+xt)^1} dt$$

$$= \frac{1}{\Gamma(b)} \int_0^1 \frac{t^{b-1}}{1+xt} dt \quad (\text{A.1})$$

However, in this expression for the case of $b < 1$ the numerator becomes like $1/t^{1-b}$ and thus the integration bound at $t = 0$ becomes divergent, which is non-ideal for quadrature. Instead one can make the variable substitution $u = t^b$ (assuming $b \neq 0$) to get:

$${}_2F_1(1, b; b+1; -x) = \frac{1}{b\Gamma(b)} \int_0^1 \frac{1}{1+xu^{1/b}} du \quad (\text{A.2})$$

which is well-behaved over the entire integration domain.

B. Proof of Large-x Asymptotic Relation

Starting from the relation (DLMF 8.17.7 and 8.17.1):

$${}_2F_1(1, b; b+1; -x) = b(-x)^{-b} \int_0^{-x} \frac{t^{b-1}}{1-t} dt \quad (\text{A.3})$$

one makes the substitution $t = -y$, $y(t = 0) = 0$, $y(t = -x) = x$, $dt/dy = -1$ to get:

$${}_2F_1(1, b; b+1; -x) = bx^b \int_0^x \frac{y^{b-1}}{1+y} dy \quad (\text{A.4})$$

Then one can take as inspiration one of the multiple equivalent definitions of the (complete) Beta function (DLMF 5.12.3):

$$B(z_1, z_2) = \int_0^\infty t^{z_1-1} (1+t)^{-(z_1+z_2)} dt \quad (\text{A.5})$$

and given the goal to establish a “large- x ” relation one can then consider the complete beta function then as the limiting case $x \rightarrow \infty$ and rewrite the original integral A.4 in terms of the complete beta function minus a small correction:

$${}_2F_1(1, b; b+1; -x) = bx^{-b} \left(\int_0^\infty \frac{y^{b-1}}{1+y} dy - \int_x^\infty \frac{y^{b-1}}{1+y} dy \right) \quad (\text{A.6})$$

Recognizing that the first term is then $B(b, 1-b)$ which has the known-relation (DLMF 8.17.3 and DLMF 5.5.1) (when coupled with the fact that $\Gamma(1) = 1$)

$$B(b, 1-b) = \frac{\Gamma(b)\Gamma(1-b)}{\Gamma(b+(1-b))} = \Gamma(b)\Gamma(1-b) = \frac{\pi}{\sin(\pi b)} \quad (\text{A.7})$$

one then arrives at the expression

$${}_2F_1(1, b; b+1; -x) = bx^{-b} \left(\frac{\pi}{\sin(\pi b)} - \int_x^\infty \frac{y^{b-1}}{1+y} dy \right) \quad (\text{A.8})$$

and the only task that remains is to give a concrete form to the second term under the assumption that x is large. One can note that an implication of the lower limit of that integral being "x" is that the integrand, y , is very large (and thus $1/y$ is very small) for all values in the integration domain and we can thus make the following power series expansion (DLMF 4.6.7) with the knowledge that it should converge rapidly:

$$\begin{aligned} \frac{y^{b-1}}{1+y} &= y^{b-1} \left(\frac{1}{y} \frac{1}{1 + \left(\frac{1}{y}\right)} \right) = y^{b-2} \left(\frac{1}{1 + \left(\frac{1}{y}\right)} \right) \\ &= y^{b-2} \left(1 - \left(\frac{1}{y}\right) + \frac{1}{2!} \left(\frac{1}{y}\right)^2 - \left(\frac{1}{y}\right)^3 + \dots \right) \quad (\text{A.9}) \end{aligned}$$

Thus the integral A.8 can be written as :

$$\begin{aligned} \int_x^\infty \frac{y^{b-1}}{1+y} dy &= \int_x^\infty \sum_{n=0}^\infty (-1)^{n+1} y^{b-2-n} dy \\ &= \sum_{n=0}^\infty (-1)^n \frac{x^{b-1-n}}{b-1-n} \quad (\text{A.10}) \end{aligned}$$

Note, that in the last step, where the integration and summation are permuted and the integral of a polynomial performed, it is assumed that term coming from the upper integral bound, x^{b-1-n} ($x \rightarrow \infty$), is for all values of n not infinite and in fact zero. This is only true if $b-1-(0) < 0$ or $b < 1$. Since $b = T/T_0$ this is not generally the case. Thus, we must expand our strategy as follows: if $b < 1$ we can immediately evaluate the hypergeometric function by truncating this series after a few terms resulting in the final expression for ${}_2F_1(1, b; b+1; -x)$ (iff $0 < b < 1$):

$$\begin{aligned} &\approx bx^{-b} \left(\frac{\pi}{\sin(\pi b)} - \frac{x^{b-1}}{b-1} + \frac{x^{b-2}}{b-2} - \frac{x^{b-3}}{b-3} \right) \\ &\approx \frac{b\pi}{\sin(\pi b)} x^{-b} - \frac{b}{b-1} x^{-1} + \frac{b}{b-2} x^{-2} - \frac{b}{b-3} x^{-3} \quad (\text{A.11}) \end{aligned}$$

and for the case of $b > 1$ (since $b = T/T_0$ a negative value never occurs) we must seek a recurrence relation that establishes an equality between ${}_2F_1(1, b; b+1; -x)$ and ${}_2F_1(1, b'-1; (b'-1)+1; -x)$. With such a relation we can then "fold" or transform the case of $b > 1$ into one where $b < 1$ and then use A.11.

C. Proof of Recurrence Relation

Beginning with the well-established contiguous relation for hypergeometric functions DLMF 15.5.16_5:

$${}_2F_1(a, b; c; z) = {}_2F_1(a-1, b; c; z) + z \frac{b}{c} {}_2F_1(a, b+1; c+1; z) \quad (\text{DLMF 15.5.16}_5)$$

Specializing to the bandtails case where $a = 1$ (thus $a-1 = 0$) and thus ${}_2F_1(0, b; c; z) = 1$ and $c = b+1$ this becomes:

$${}_2F_1(1, b; b+1; z) = 1 + z \frac{b}{b+1} {}_2F_1(1, b+1; (b+1)+1; z) \quad (\text{A.12})$$

Making the substitution $b+1 \rightarrow b$:

$${}_2F_1(1, b-1; (b-1)+1; z) = 1 + z \frac{b-1}{b} {}_2F_1(1, b; b+1; z) \quad (\text{A.13})$$

and then rearranging one gets the following final result that allows one to take any case where $b > 1$ and recurse it untily $b < 1$:

$${}_2F_1(1, b; b+1; z) = \frac{b}{b-1} z^{-1} ({}_2F_1(1, b-1; (b-1)+1; z) - 1) \quad (\text{A.14})$$

D. Proof of Derivative Relationship

Starting from the relation (combining DLMF 8.17.7 and 8.17.1):

$${}_2F_1(1, b; b+1; -x) = bx^{-b} \int_0^x \frac{y^{b-1}}{1+y} dy \quad (\text{A.15})$$

and taking the derivative one gets:

$$\begin{aligned} \frac{d}{d(-x)} {}_2F_1(1, b; b+1; -x) &= \\ &-b \left(\frac{d}{dx} x^{-b} \right) \int_0^x \frac{y^{b-1}}{1+y} dy - bx^{-b} \left(\frac{d}{dx} \int_0^x \frac{y^{b-1}}{1+y} dy \right) \quad (\text{A.16}) \end{aligned}$$

The first here term is:

$$\begin{aligned} b^2 x^{-b-1} \int_0^x \frac{y^{b-1}}{1+y} dy &= \frac{b}{x} \left(bx^{-b} \int_0^x \frac{y^{b-1}}{1+y} dy \right) \\ &= -\frac{b}{(-x)} {}_2F_1(1, b; b+1; -x) \quad (\text{A.17}) \end{aligned}$$

and the second term is (using Leibniz' rule):

$$-bx^{-b} \left(\frac{d}{dx} \int_0^x \frac{y^{b-1}}{1+y} dy \right) = -bx^{-b} \left(\frac{x^{b-1}}{1+x} \right) = -\frac{b}{x(1+x)} \quad (\text{A.18})$$

Thus one arrives at the following final expression which allows the derivative to be evaluated by re-using the value of ${}_2F_1$ that was already calculated:

$$\frac{d}{dz} {}_2F_1(1, b; b+1; z) = -\frac{b}{z} {}_2F_1(1, b; b+1; z) + \frac{b}{z(1-z)} \quad (\text{A.19})$$

REFERENCES

- [1] A. Beckers, et al. arXiv:1811.09146.,
- [2] M. Thesberg, et al. Micromachines 15.7 (2024): 829.
- [3] A. Beckers, et al., J. Appl. Phys. 129.4 (2021).
- [4] Boost C++ Libraries (1.72.0 (2019) or later). <http://www.boost.org/>
- [5] NIST Digital Library of Mathematical Functions. <https://dlmf.nist.gov/>
- [6] Minimos-NT. [Online]. Available: <https://globaltcad.com/minimos-nt>
- [7] Z. Stanojević, et al., IEEE Trans. Elect. Dev. 2022, 70, 708–713
- [8] S. Jin, et al., SISPAD 2021 (pp. 251-254).