# Numerical pathology in selected Kubelka-Munk formulas, and strategies for mitigation 

J A Stephen Viggiano, PhD

19 May 2021

Causes of numerical pathology in formulas for reflectance factor $(R)$, transmittance factor $(T)$, and reflectance factor over a perfectly black background $\left(R_{0}\right)$ under the Kubelka-Munk model are posited, and alternate formulas believed less prone to these pathologies are introduced. Suggestions are offered not only for $R, T$, and $R_{0}$, but also for intermediate or adjunct quantities used in the main formulas. Computational experiments were performed to verify the new models produce the same results as the existing ones under non-pathological conditions, exhibit acceptable levels of precision in a customary floating-point environment, and are more robust with respect to edge cases where an input quantity is zero. The new formulas performed well, with some evidence that the new hyperbolic forms provide better accuracy than their exponential counterparts.

Keywords: Kubelka-Munk, numerical pathology.

## Symbols and Notation

$R$ Reflectance factor, particularly of a colorant layer in optical contact with a backing
$R_{\infty}$ Reflectivity, i.e., the reflectance factor of a colorant layer so thick that its reflectance factor is independent of the background
$R_{g}$ Reflectance factor of the backing or substrate
$R_{0}$ Reflectance factor of the colorant layer over a perfectly absorbing background ( $R$, when $\left.R_{g}=0\right)$
$T$ Transmittance factor (here, of a colorant layer)
$X$ Thickness of the colorant layer
$K$ Coefficient of absorption of the colorant; the proportion of light absorbed in a layer of infinitesimal thickness $d x$ will be $K d x$.
$S$ Coefficient of scatter of the colorant; the proportion of light scattered in a layer of infinitesimal thickness $d x$ will be $S d x$.
$a$ Used in so-called hyperbolic solutions; $a=\frac{K+S}{S}$
$b$ Also used in hyperbolic solutions; $a^{2}-b^{2}=1$, so $b=\sqrt{a^{2}-1}$
$L=\sqrt{(K+S)^{2}-S^{2}}=\sqrt{K(K+2 S)}$
$M=K+S$
$P=L+K+S=L+M$

## Notes

The notation used in this paper follows, wherever possible, that used by Kubelka. [1] $L$ is from Gurevič. [2] The last two quantities, $M$ and $P$, are introduced here for convenience.

Other than layer thickness, $X$, the quantities are dependent on wavelength, and, by the first law of thermodynamics, [3, p 130] cannot be negative, nor can $a$ be less than one.

The units for layer thickness are customarily micrometers. However, the author has on occasion used grams per square meter as a convenient surrogate for thickness; it may be translated into thickness if the mass density of the colorant is known.

The units for $K, S, L, M$, and $P$ are the reciprocal of the units used for thickness, e.g., reciprocal micrometers.

All of the other quantities are dimensionless.

## 1 Introduction

Kubelka and Munk investigated optical properties of homogeneous layers using a system of two first-order linear differential equations, one equation each for two fluxes, each traversing a layer a layer illuminated on one side in two directions: the first away from the illuminated surface, and the second back toward the illuminated surface.[4] There are several excellent references that provide an introduction to Kubelka-Munk theory, including Wyszecki and Stiles, Judd and Wyszecki. Philips-Invernizzi, et al.,[5] provide a thorough literature review with historical and modern context. Berns provides an approachable introduction.

The Kubelka-Munk model has been applied to a wide variety of problems, including colorant mixing [6, 7, 8], astrophysics, remote sensing, and color hardcopy. [9]

In this paper, formulas derived from the Kubelka-Munk model (and Gurevičs [2] nearly identical model), will be examined to identify, and, hopefully mollify, issues that may arise during machine computation. The formulas are to compute $R$, the reflectance factor of a col-
orant layer atop a background of known reflectance factor; $T$, the transmittance factor of a colorant layer; and $R_{0}$, the reflectance factor of a colorant layer atop a perfectly black backing.

### 1.1 Computational pathology in formulas for Kubelka-Munk models

There are two main computational pathologies examined in this paper: computational failures, and excessive loss of precision.

A computational failure is an error at run time that results in a crash, an exception being raised (which may result in a crash), or a "Not-A-Number" (NaN) result. The specific computational failures most frequently encountered by the author in work with Kubelka-Munk models have been:

- attempting to take the hyperbolic cotangent of 0 ; and
- attempting to divide by 0 or take the reciprocal of 0 .

Kubelka clearly stated that the hyperbolic forms were intended for hand computation. Even when using a calculating machine, the human operating it could quickly detect a fault and apply an alternate formula; Kubelka provided several of these in his 1948 paper. Anticipating and catching all fault-inducing situations, and handling them appropriately, is not as simple when the calculations are performed by a computer program.

A second form of computational pathology is excessive loss of precision. Floating-point numbers are stored with a finite amount of precision, usually either 24 or 53 bits of binary precision. These translate into approximately seven and 16 decimal digits of precision, respectively.

Precision is lost nearly every time a floating-point operation is performed. In particular, a large amount of precision may be lost when two floating point numbers close in value are subtracted. [10, 11] In Figure 1, two floating-point values in IEEE-724 half-precision are subtracted. In this format, there are 12 digits of binary precision, or approximately 3.6 decimal
digits of precision. The difference, unfortunately, has only five significant binary digits (approximately 1.5 significant decimal digits), meaning a loss of seven binary digits, or more than 2 decimal digits.

The seven decimal digits offered by single precision, and even the nearly 16 digits afforded by double precision, can be quickly eroded by injudicious arithmetic operations. Unlike computational failures, precision loss can nuanced and difficult to detect.

### 1.2 Problem, goals, and approach

To succinctly state the problem, pathological results, including faults and loss of precision, may arise when performing calculations using the Kubelka-Munk model. These will be particularly acute with machine computation.

The goals of this paper, with respect to formulas based on the Kubelka-Munk model, are:

- Identify pathological calculations resulting in faults and excessive precision loss.
- When more than one formula is available, determine which is less prone to these pathologies.
- Rule out possible suspected causes of pathology.
- Offer alternate formulas less prone to faults and/or excessive loss of precision.
- In short, establish a nucleus of best practices for machine computation.

The approach may be gleaned from the organization of this paper. In the next section, existing formulas will be examined for potential causes of pathology. In the following section, identities that have potential to reduce computational pathologies will be derived. Next, alternate formulas will be evaluated in customary and high-precision computational environments. Finally, conclusions and recommendations for best practices will be offered.

## 2 Formulas in the prior art

### 2.1 Formulas for reflectance

### 2.1.1 Kubelka and Munk, 1931 [4]

$$
\begin{equation*}
R=\frac{\left(R_{g}-R_{\infty}\right) / R_{\infty}-R_{\infty}\left(R_{g}-1 / R_{\infty}\right) \exp \left[S X\left(1 / R_{\infty}-R_{\infty}\right)\right]}{R_{g}-R_{\infty}-\left(R_{g}-1 / R_{\infty}\right) \exp \left[S X\left(1 / R_{\infty}-R_{\infty}\right)\right]} \tag{1}
\end{equation*}
$$

This formula contains four different subtractions, and the reciprocal of $R_{\infty}$ appears in several places. $R_{\infty}$ will approach zero as the colorant approaches perfect transparency (i.e., as $S \rightarrow 0$ ).

While this may appear rather complicated, there are other ways to express it in less complicated terms. One follows immediately below, three more are offered in a later section of the paper.

### 2.1.2 Kubelka, 1948 [1]

Kubelka's so-called hyperbolic solution,

$$
\begin{equation*}
R=\frac{1-R_{g}(a-b \cdot \operatorname{coth} b S X)}{a-R_{g}+b \cdot \operatorname{coth} b S X}, \tag{2}
\end{equation*}
$$

does appear simpler than Eq (1), and involves two fewer subtractions. Unfortunately, it contains the hyperbolic cotangent, which is singular at zero, as shown in Figure 2. While not a problem for hand computation, this is best avoided with machine computation. When any of $b$, $S$, or $X$ are zero, an arithmetical error will occur.

### 2.2 Formulas for transmittance

### 2.2.1 Gurevič, 1930 [2, p 756]

Kubelka and Munk did not address transmittance in their 1931 paper. Gurevič, in a paper published slightly earlier than Kubelka and Munk's, did provide an expression for transmittance.

Translating Gurevič's notation ${ }^{1}$ to that of the present paper, it was:

$$
\begin{equation*}
T=\frac{\left(1-R_{\infty}^{2}\right) \exp (-L X)}{1-R_{\infty}^{2} \exp (-2 L X)} \tag{3}
\end{equation*}
$$

where:

$$
\begin{equation*}
L=\sqrt{(K+S)^{2}-S^{2}} . \tag{4}
\end{equation*}
$$

In a later section, $L$ will be written in a more numerically favorable form, further interpreted in terms of the Kubelka-Munk framework, and employed in formulas for machine computation.

### 2.2.2 Kubelka, 1948 [1]

Kubelka offered a hyperbolic solution for transmittance:

$$
\begin{equation*}
T=\frac{b}{a \sinh b S X+b \cosh b S X} \tag{5}
\end{equation*}
$$

This is susceptible to instability if $b$ is small, which it will be for white media. A very small value of $b, S$, or $X$ can make the first term in the denominator effectively vanish through floating point underflow because $\sinh (0)=0$. Further, because it is a factor in the second term, it can likewise cause this term to become pathologically small. A singularity will of course result if both terms in the denominator vanish.

### 2.3 Formulas for $R_{0}$

While $R_{0}$ is a special case of reflectance, Kubelka and Munk[4] disclosed the following formula (their $\mathrm{Eq}(6))$ for $R_{0}$, the reflectance of a colorant layer over a perfectly black substrate. In the

[^0]notation used in the present paper, it was:
\[

$$
\begin{equation*}
R_{0}=\frac{\exp \left[\left(1 / R_{\infty}-R_{\infty}\right) S X\right]-1}{1 / R_{\infty} \exp \left[\left(1 / R_{\infty}-R_{\infty}\right) S X\right]-R_{\infty}} \tag{6}
\end{equation*}
$$

\]

and may be easily derived from Eq (1) by substituting 0 in it in place of $R_{g}$.
Kubelka [1] also offered two related hyperbolic expressions. The first,

$$
\begin{equation*}
R_{0}=\frac{1}{a+b \operatorname{coth} b S X} \tag{7}
\end{equation*}
$$

is easily obtained from Eq (2); his second may be obtained by multiplying both numerator and denominator of Eq (7) by $\sinh b S X$ :

$$
\begin{equation*}
R_{0}=\frac{\sinh b S X}{a \sinh b S X+b \cosh b S X} \tag{7a}
\end{equation*}
$$

Eq (6) does have the reciprocal of $R_{\infty}$, which is pathological as $S \rightarrow 0$. As the lead factor in the first term of the denominator, it is easily remedied. The instances in the two exponentials, however, will not yield to this simple expedient.

The pathology of Eq (7) is identical to that of the hyperbolic form for reflectance given in Eq (2). Inasmuch as their denominators are identical, the potential pathologies of Eq (5) and Eq (7a) are likewise the same.

## 3 Remediation of pathologies

Kubelka [1] discussed several approximate formulas for $R, T, R_{0}$, and related quantities. While performing hand-computation, certain edge cases that might cause problems could be avoided, and the calculation could be simplified. Kubelka's Table III, and the text that introduces it, elegantly address several special cases through 28 simplified formulas.

For machine computation, the general case/special case approach championed and enabled by Kubelka may not be the best strategy. First, very specific boundaries between the general case and each of the special cases are required. Logic and branching to handle the special cases complicate coding and will require a much larger testing and debugging effort. Further, with machine computation, discontinuities and other inconsistencies as calculation switches between general case and one of the special cases are undesirable, and, hopefully may be avoided. Enabling robust computation using general-case formulas for all cases, eliminating the need to identify, code, and test edge cases, is the goal of this paper.

### 3.1 Enabling lemmatic foundations

In the following section, strategies for more robust machine computation of $R, T$, and $R_{0}$ under the Kubelka-Munk model will be derived and discussed, using some enabling lemmas presented here.

Recall Gurevič's $L$ (Eq (4)); we avoid a subtraction if we write instead:

$$
L=\frac{\sqrt{K(K+2 S)}}{S}
$$

Two convenient forms for $R_{\infty}$ are worth knowing:

$$
R_{\infty}=a-b=\frac{1}{a+b} .
$$

By definition, $a=(K+S) \div S$, and $b=\sqrt{a^{2}-1}$. Another way to represent $b$ appears several times in the literature, and is easily derived from these two definitions:

$$
b=\sqrt{\frac{K}{S}\left(2+\frac{K}{S}\right)}
$$

This may be further simplified to:

$$
b=\frac{\sqrt{K(K+2 S)}}{S}=\frac{L}{S}
$$

Therefore, an equivalent representation for the germ of some pathology discussed earlier is:

$$
b S=L,
$$

While this identity applies regardless of the value of $S$, there is a limit of interest:

$$
\lim _{S \rightarrow 0} b S X=K X
$$

(A side note: As scattering vanishes, the argument to the transcendental functions in many of the formulas in Kubelka-Munk theory approaches the product $K X$. The absorption coefficient, $K$, plays a similar role here to the extinction coefficient, $\varepsilon$, in the Bouguer-Lambert-Beer [12, 13, 14] cannon, where it multiplies the concentration and path length to form the argument to the exponential function.)

Another useful representation of $b$ is:

$$
b=\frac{1}{2}\left(\frac{1}{R_{\infty}}-R_{\infty}\right)
$$

whence:

$$
\frac{1}{R_{\infty}}-R_{\infty}=2 b
$$

### 3.1.1 Preferred forms for $R_{\infty}$ and related quantities

To avoid cancellation of precision caused by subtraction, the following formulas are recommended:

$$
\begin{align*}
R_{\infty} & =\frac{S}{L+M}=\frac{S}{P}  \tag{8}\\
1-R_{\infty} & =\frac{K+L}{L+M}=\frac{K+L}{P} \tag{9}
\end{align*}
$$

where $P=L+M=K+S+L$.
Gurevič's formula for transmittance contains the factor $1-R_{\infty}^{2}$. This may be factored into $\left(1+R_{\infty}\right)\left(1-R_{\infty}\right)$, which may be written:

$$
\begin{equation*}
1-R_{\infty}^{2}=\frac{(P+S)(K+L)}{P^{2}} \tag{10}
\end{equation*}
$$

## 4 Revised formulas

It should be noted that the following formulas are mathematically equivalent to the corresponding formula appearing above, and produce identical results under non-pathological conditions, as will be shown in the following section. The equation numbers in the following revised formulas are based on the number of the corresponding equation from earlier, with a "-R" (and possibly an additional letter where two or more alternatives are offered).

### 4.1 Alternate formulas for $R$

### 4.1.1 Exponential form

Making substitutions using identities derived in the previous section, Kubelka and Munk's 1931 formula for reflectance may be cast as:

$$
R=\frac{P\left(P R_{g}-S\right)+S\left(P-S R_{g}\right) \exp (2 L X)}{S\left(P R_{g}-S\right)+P\left(P-S R_{g}\right) \exp (2 L X)}
$$

Note that the following term is repeated (for convenience, associated here with the symbol $\alpha$ ):

$$
\alpha=\left(P-S R_{g}\right) \exp (2 L X)
$$

The revised formula may now be written:

$$
\begin{equation*}
R=\frac{P\left(P R_{g}-S\right)+S \alpha}{S\left(P R_{g}-S\right)+P \alpha} \tag{1-R}
\end{equation*}
$$

Discussion. This last form appears to reduce the occasion of the pitfalls identified, namely, loss of floating-point precision and/or floating-point failure as $X \rightarrow 0, R_{\infty} \rightarrow 0$, and either $S \rightarrow 0$ or $K \rightarrow 0$. This form has just two distinct subtractions, vis-à-vis the four required in Eq (1), and, in contrast to the hyperbolic form, no issues as $X \rightarrow 0$.

Also, note that for perfect transparency, with $K>S=0, R_{\infty}=0, L=K$, and $P=2 K$, so the numerator will equal $4 K^{2} R_{g}$, while the denominator will be $4 K^{2} \exp (2 K X)$, yielding $R=R_{g} \exp (-2 K X)$, which is in direct concordance with the Bouguer-Lambert-Beer model.

Aside from the two subtractions, one pathology remains, for $K=0$. Then, $L=0, P=S$, $\alpha=S\left(1-R_{g}\right)$, and the denominator becomes zero.

### 4.1.2 Hyperbolic forms

Two alternatives that avoid the issues with the hyperbolic cotangent, but nevertheless employ hyperbolic functions, are:

$$
\begin{equation*}
R=\frac{L \cdot R_{g}+\left(S-M \cdot R_{g}\right) \tanh L X}{L+\left(M-S \cdot R_{g}\right) \tanh L X} \tag{2-R}
\end{equation*}
$$

and

$$
\begin{equation*}
R=\frac{L \cdot R_{g} \cosh L X+\left(S-M \cdot R_{g}\right) \sinh L X}{L \cdot \cosh L X+\left(M-S \cdot R_{g}\right) \sinh L X} . \tag{2-Ra}
\end{equation*}
$$

These are both less susceptible to numerical pathology as $X \rightarrow 0$, and contain one fewer subtraction than the original hyperbolic form in Eq (2), in addition to the implicit subtraction eliminated by no longer using the quantity $b$.

### 4.2 Formulas for $T$

### 4.2.1 Revision of Gurevič's formula for transmittance

A subtraction is eliminated by using the preferred form of $1-R_{\infty}$, factoring $1-R_{\infty}^{2}$, and :

$$
\begin{equation*}
T=\frac{(P+S)(K+L) \exp (-L X)}{P^{2}-S^{2} \exp (-2 L X)} \tag{3-R}
\end{equation*}
$$

When the input is $K, S$, and $X$, only one subtraction is used, and appears to be stable except when $K=0$, resulting in a $0 / 0$ indeterminate form.

### 4.2.2 Hyperbolic form

Making the substitution $L=b S$ in Eq (5), and multiplying numerator and denominator by $S$, one obtains:

$$
\begin{equation*}
T=\frac{L}{M \sinh L X+L \cosh L X} \tag{5-R}
\end{equation*}
$$

Other than the division by two and subtraction implicit in both hyperbolic functions, only one division, and no subtractions are employed by this form. The denominator will vanish only as both $K$ and $S$ do so.

### 4.3 Formulas for $R_{0}$

### 4.3.1 Revised exponential form

$$
\begin{equation*}
R_{0}=\frac{S P[\exp (2 L X)-1]}{P^{2} \exp (2 L X)-S^{2}} \tag{6-R}
\end{equation*}
$$

### 4.3.2 Revised hyperbolic forms

$$
\begin{gather*}
R_{0}=\frac{S \tanh L X}{L+M \tanh L X},  \tag{7-R}\\
R_{0}=\frac{S \sinh L X}{M \sinh L X+L \cosh L X} \tag{7a-R}
\end{gather*}
$$

## 5 Experimental

Computational experiments were performed to assess agreement among the formulas, evaluate precision under customary floating-point conditions, and check behavior under extreme cases.

### 5.1 Part 1: Verification

In order to verify that the revised formulas produce the same results as the prior forms under most conditions (i.e., other than $K=0$ or $X=0$ ), the original and revised formulas were coded in the C++ programming language. In order to rule out all but the smallest rounding errors as cause for any discrepancy, special arithmetic was employed. The GNU MPFR package [15]
was used, providing over 1600 binary digits (at least 500 decimal digits) of precision. The Boost Multiprecision wrapper [16] provided a convenient interface for the MPFR capabilities.

Details of the testing platform appear in Table 1.
Fifty thousand random combinations of $K, S, X$, and $R_{g}$ were generated using the default random number generator in the GNU g++ environment on a GNU/Linux workstation. The random number generator was initialized with a constant seed to ensure consistent results from one run to the next. The distributions of all four variables were uniform, with $K, S$, and $X$ uniformly distributed on $[0,2.5]$, while $R_{g}$ was drawn from a uniform distribution on $[0,1]$. The values generated by each combination of the two existing and three new reflectance formulas, two existing and two new transmittance formulas, and three existing and three new formulas were compared on a pairwise basis.

For reflectance, the formulas compared were those in Equations (1), (2), (1-R), (2-R), and (2a-R). The transmittance formulas compared were those in Equations (3), (5), (3-R), and (5R). Finally, the formulas for $R_{0}$ compared were given in Equations (6), (7), (7a), (6-R), (7-R), and ( $7 \mathrm{a}-\mathrm{R}$ ).

The maximum absolute difference for each pair appear in Table 2. For example, the reflectances calculated using the five formulas identified in the previous paragraph differed from each other by no more than $1.207 \times 10^{-499}$ when calculated in a floating-point environment with 500 decimal digits of precision. With maximum differences on the order of the floatingpoint epsilon, there is assurance that the old and new formulas produce, in essence, identical results under non-pathogenic conditions.

### 5.2 Part 2: Evaluation of floating-point results

The same models were exercised, using the same combinations of $K, S, X$, and $R_{g}$, but with the computations performed in double precision ( 53 bits of binary precision) as well. This environment represents practical computation conditions. Rather than the disagreement between
the formulas, the maximum absolute value by which the customary precision calculations differ from the high-precision counterparts were computed. In earlier experiment it was shown that the formulas for each mode agree, so, for each set of input values ( $K, S, X$, and, for reflectance, $R_{g}$ ) the high-precision answers were averaged and regarded as the true value for that input set. The absolute value of the difference between this "truth" and an answer computed using the customary precision level was taken as an error. The maximum errors for the 50000 trials appear in Table 3.

These results show agreement more than adequate for practical calculations, and are consistent with small floating point errors.

### 5.3 Part 3: Evaluation under pathogenic conditions

An additional run was made with $K=0$, another with $S=0$, and a third with $X=0$. A fourth was made for reflectance with $R g=0$.

As anticipated, all formulas generated a not a number result with $K=0$. The complete set of results appear in Table 4.

### 5.4 Discussion

In spite of the author's predictions, the existing formulas performed very well, except when $K=0$ (for which all formulas failed), and other pathogenic conditions. The new formulas produced results that were reasonable for all other pathogenic conditions, as shown in Table 4, and agreement between high and customary precision, as shown in Table 3 provide assurance that their answers are essentially correct.

The results shown in Table 3 indicate that the hyperbolic formulas have an edge over the exponential forms, particularly for $R_{0}$.

## 6 Conclusions

In general, the new formulas performed as well as their existing counterparts. The revised exponential forms showed only slighter worse accuracy, while the hyperbolic forms showed slighter better accuracy.

Those starting new projects should consider adopting the new formulas, in particular, those using hyperbolic functions, as well as improved formulas for $b S, R_{\infty}$, and other adjunct quantities. Even those that wish to continue to use the existing formulas can profit from new formulas for adjunct quantities, with fewer subtractions and denominators less likely to vanish. However, with the new formulas producing fewer "Not-a-Number" results under the pathogenic conditions, it may be worthwhile to at least consider including them in existing projects.

The quantity $K+S$, called $M$ in this paper, plays a similar role to $a$ in hyperbolic solutions. Likewise, $L$ is an analog to $b$. This is not surprising, because $a=M \div S$, and $b=L \div S$.

The Bouguer-Lambert-Beer model may be more easily understood as a special case of KubelkaMunk, because forms introduced here more easily handle vanishing scatter and segue gracefully to formulas based on Bouguer-Lambert-Beer.

## References

[1] Paul Kubelka. New contributions to the optics of intensely light-scattering materials. Part I. Journal of the Optical Society of America, 38(5):448-458, 1948.
[2] M Gurevič. Über eine rationelle Klassifikation der lichtenstreuenden Medien. Physikalische Zeitschrift, 31:753-763, 1930.
[3] Alan M Whitman. Thermodynamics: Basic Principles and Engineering Applications. Springer, Cham, 2020.
[4] Paul Kubelka and Franz Munk. Ein Beitrag zur Optik der Farbenstriche. Zeitschrift für Technische Physik, 12:593-601, 1931.
[5] Bernadette Philips-Invernizzi, Daniel Dupont, and Claude Cazé. Bibliographical review for reflectance of diffusing media. Optical Engineering, 40(6):1082-1092, 2001.
[6] D R Duncan. The colour of pigment mixtures. Proceedings of the Physical Society, 52:390-401, 1940.
[7] Eugene Allen. Basic equations used in computer color matching. Journal of the Optical Society of America, 56(9):1256-1259, 1966.
[8] Eugene Allen. Basic equations used in computer color matching, II. Tristimulus match, two-constant theory. Journal of the Optical Society of America, 64(7):991-993, 1974.
[9] J A Stephen Viggiano. New Models for the Reflectance Spectra of Halftone-based Color Hardcopy. PhD dissertation, Rochester Institute of Technology, Rochester, NY, USA, 2010.
[10] Walter Gautschi. Machine arithmetic. In Numerical Analysis, pages 1-54. Birkhäuser, Boston, MA, 2012.
[11] Petre Teodorescu, Nicolae-Doru Stanescu, and Nicolae Pandrea. Errors in numerical analysis. In Numerical Analysis with Applications in Mechanics and Engineering, pages 1-15. Wiley/IEEE Press, Hoboken, NJ, 2013.
[12] [Pierre] Bouguer. Essai d'Optique sur la Gradation de la Lumière. Claude Jombert, Paris, 1729. Section III, Probleme III-IV, Pages 73-81.
[13] J H Lambert. Photometria, sive, de mensura et gradibus luminus, colorum et umbrae. Eberhard Klett, 1760.
[14] A Beer. Bestimmung der Absorption des rothen Lichts in farbingen Flüssigkeiten. Annalen Physik und Chemie, 86(2):78-90, 1852.
[15] Laurent Fousse, Guillaume Hanrot, Vincent Lefèvre, Patrick Pélissier, and Paul Zimmermann. MPFR: A multiple-precision binary floating-point library with correct rounding. ACM Transactions on Mathematical Software, 33(2):13.1-13.15, June 2007.
[16] John Maddock and Christopher Kormanyos. Chapter 1. Boost.Multiprecision. https://www.boost.org/doc/libs/1_76_0/libs/multiprecision/doc/html/index.html. Accessed 2021-05-28.

12 significant binary digits in;

$$
\begin{array}{r}
1.00001110001 \times 2^{4} \\
-1.00001010010 \times 2^{4} \\
0.00000011111 \times 2^{4} \\
=1.11110000000 \times 2^{-3}
\end{array}
$$

5 significant binary digits out
Figure 1: Example of loss of precision.


Figure 2: Graph of the hyperbolic cotangent function in the vicinity of zero.

Table 1: Details of testing platform.

Manufacturer
Model
Processor
Topology
Memory
Operating System
Compiler
Customary precision
High precision

Lenovo 20KH002FUS
Intel Core i7-8650U, 4.20 GHz
1 Processor, 4 Cores, 8 Threads
15.4 GB
x86_64 GNU/Linux, kernel 5.8.0
GNU g++, version 9.3.0
g++ C++ double (IEEE 754 float64)
GNU MPFR, version 4.0.2, 500 decimal digits

Table 2: Comparison of formulas, high precision

| Quantity | Maximum Difference |
| :---: | :---: |
| $R$ | $1.207 \times 10^{-499}$ |
| $T$ | $1.6582 \times 10^{-499}$ |
| $R_{0}$ | $9.3836 \times 10^{-500}$ |

Maximum formula-to-formula difference for 50000 random combinations of $K, S, X$, and $R_{g}$, computed in a floating-point environment providing at least 500 decimal digits of precision.
Table 3: High- vs customary precision results.

| Formulas for $R_{0}$ |  |
| :---: | :---: |
| $\mathrm{Eq}(6)$ | $2.6879 \times 10^{-15}$ |
| $\mathrm{Eq}(7)$ | $1.0293 \times 10^{-26}$ |
| $\mathrm{Eq} \mathrm{(7a)}$ | $5.5957 \times 10^{-27}$ |
| $\mathrm{Eq}(6-\mathrm{R})$ | $2.6879 \times 10^{-15}$ |
| $\mathrm{Eq}(7-\mathrm{R})$ | $8.3083 \times 10^{-27}$ |
| $\mathrm{Eq}(7 \mathrm{a}-\mathrm{R})$ | $6.8277 \times 10^{-27}$ |

Equation numbers containing "-R" are revisions introduced in this paper. Values are largest absolute value difference between quantities computed in a 500-decimal digit environment and a customary 17 decimal double-precision environment for 50000 random combinations of input.

Table 4: Results using pathological input.

| K | Input |  |  |  | Table 4a: Formulas for $R$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $S$ | $X$ | $R_{g}$ | Eq (1) | Eq (2) | Eq ( $1-\mathrm{R}$ ) | Eq (2-R) |
| 0 | 1 | 1 | 0.85 | nan | nan | nan | nan |
| 1 | 0 | 1 | 0.85 | nan | nan | 0.11503 | 0.11503 |
| 1 | 1 | 0 | 0.85 | 0.8500 | nan | 0.8500 | 0.8500 |
| 1 | 1 | 1 | 0.00 | 0.26015 | 0.26015 | 0.26015 | 0.26015 |
|  | Input |  |  | Table 4b: Formulas for $T$ |  |  |  |
|  | K | S | X | Eq (3) | Eq (5) | Eq (3-R) | Eq (5-R) |
|  | 0 | 1 | 1 | nan | nan | nan | nan |
|  | 1 | 0 | 1 | nan | nan | 0.36788 | 0.36788 |
|  | 1 | 1 | 0 | 1 | 1 | 1 | 1 |


| Input |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $K$ | $S$ | $X$ | $\mathrm{Eq}(6)$ | $\mathrm{Eq}(7)$ | $\mathrm{Eq}(7 \mathrm{a})$ | $\mathrm{Eq}(6-\mathrm{R})$ | $\mathrm{Eq}(7-\mathrm{R})$ | $\mathrm{Eq}(7 \mathrm{a}-\mathrm{R})$ |
| 0 | 1 | 1 | nan | nan | nan | nan | nan | nan |
| 1 | 0 | 1 | nan | nan | nan | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Equation numbers containing "- R " are revisions introduced in this paper.
"nan" is "not-a-number" resulting from division by zero or similar error.


[^0]:    ${ }^{1}$ Gurevič's parameters were congruent to Kubelka and Munk's $K+S$ and $S$.

