Astronomy & Astrophysics

TRUST

I. A 3D externally illuminated slab benchmark for dust radiative transfer

K. D. Gordon[1,2], M. Baes[2], S. Bianchi[3], P. Camps[2], M. Juvela[4], R. Kuiper[5,6], T. Lunttila[8], K. A. Misselt[10], G. Natale[9], T. Robitaille[6,7], and J. Steinacker[11,12,6]

[1] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
   e-mail: kgordon@stsci.edu
[2] Sterrenkundig Observatorium, Universiteit Gent, Krijgslaan 281 S9, 9000 Gent, Belgium
[3] INAF–Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125 Firenze, Italy
[4] Department of Physics, PO Box 64, University of Helsinki, 00014 Helsinki, Finland
[5] Institut für Astronomie und Astrophysik, Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany
[6] Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
[7] Freelance Consultant, Headingley Enterprise and Arts Centre, Bennett Road Headingley, Leeds LS6 3HN, UK
[8] Dept. of Earth and Space Sciences, Chalmers Univ. of Technology, Onsala Space Observatory, 439 92 Onsala, Sweden
[9] Jeremiah Horrocks Institute, University of Central Lancashire, Preston, PR1 2HE, UK
[10] Steward Observatory, University of Arizona, Tucson, AZ 85721, USA
[11] Univ. Grenoble Alpes, IPAG, 38000 Grenoble, France
[12] CNRS, IPAG, 38000 Grenoble, France

ABSTRACT

Context. The radiative transport of photons through arbitrary three-dimensional (3D) structures of dust is a challenging problem due to the anisotropic scattering of dust grains and strong coupling between different spatial regions. The radiative transfer problem in 3D is solved using Monte Carlo or Ray Tracing techniques as no full analytic solution exists for the true 3D structures.
Aims. We provide the first 3D dust radiative transfer benchmark composed of a slab of dust with uniform density externally illuminated by a star. This simple 3D benchmark is explicitly formulated to provide tests of the different components of the radiative transfer problem including dust absorption, scattering, and emission.
Methods. The details of the external star, the slab itself, and the dust properties are provided. This benchmark includes models with a range of dust optical depths fully probing cases that are optically thin at all wavelengths to optically thick at most wavelengths. The dust properties adopted are characteristic of the diffuse Milky Way interstellar medium. This benchmark includes solutions for the full dust emission including single photon (stochastic) heating as well as two simplifying approximations: One where all grains are considered in equilibrium with the radiation field and one where the emission is from a single effective grain with size-distribution-averaged properties. A total of six Monte Carlo codes and one Ray Tracing code provide solutions to this benchmark.
Results. The solution to this benchmark is given as global spectral energy distributions (SEDs) and images at select diagnostic wavelengths from the ultraviolet through the infrared. Comparison of the results revealed that the global SEDs are consistent on average to a few percent for all but the scattered stellar flux at very high optical depths. The image results are consistent within 10%, again except for the stellar scattered flux at very high optical depths. The lack of agreement between different codes of the scattered flux at high optical depths is quantified for the first time. Convergence tests using one of the Monte Carlo codes illustrate the sensitivity of the solutions to various model parameters.
Conclusions. We provide the first 3D dust radiative transfer benchmark and validate the accuracy of this benchmark through comparisons between multiple independent codes and detailed convergence tests.

Key words. methods: numerical – ISM: general – radiative transfer

1. Introduction

The transport of radiation through dust plays a central role in many astrophysical objects as dust is efficient in absorbing and scattering ultraviolet (UV) through near-infrared photons and re-radiating the absorbed energy in the infrared and submillimeter (submm). The interpretation of observations from the UV to the submm is critically linked to an accurate calculation of the radiative transfer (RT) through dust. The dust RT problem is difficult to solve as the time-independent version is six dimensional (space, angle, and wavelength) with strongly anisotropic scattering and non-linear coupling between different spatial regions.

The complexity of the solution is especially evident when the object of interest is intrinsically three-dimensional (3D). Solutions using ray-tracing and Monte Carlo techniques (and mixtures of the two) are used in modern RT codes to solve this problem. Steinacker et al. (2013) gives an in-depth review of the 3D RT problem, current solution techniques, and an overview of existing codes, the number of which has grown significantly in the last 15 yr.

While it is common to provide benchmark solutions for specific objects to ensure that different codes produce the same answer within some tolerance, there are no existing intrinsically 3D RT benchmarks; existing benchmarks focus

Table 1. Slab setup details.

Name	Values
Dust Geometry	
System size	$10 \times 10 \times 10$ pc
System coordinates	-5 to 5 pc
Slab z extent	-2 to -5 pc
Slab xy extent	-5 to 5 pc
$\tau_z(1\,\mu m)$	0.01, 0.1, 1, 10
Blackbody source	
Location (x, y, z)	(0 pc, 0 pc, 4 pc)
Temperature	10 000 K
Luminosity	100 000 $L_\odot$
	3.839×10^{38} ergs s^{-1}
Observer	
Distance	10 000 pc

Fig. 1. Setup of the slab benchmark is graphically illustrated. The external views for the output SEDs and images are shown.

on one-dimensional (1D) or two-dimensional (2D) objects (Ivezic et al. 1997; Pascucci et al. 2004; Pinte et al. 2009). In addition to being 1D or 2D, existing RT benchmarks do not include the full dust radiative transfer solution, using approximations in either dust scattering (e.g., isotropic only) or dust emission (e.g., equilibrium only emission) processes. This has motivated a group of 3D dust RT coders to come together and propose a suite of 3D benchmarks that will test the many aspects of the dust RT solution in a range of geometries. This suite of benchmarks is named TRUST (benchmarks for the Transport of Radiation through a dUSTy medium)[1].

In general, code benchmarks are motivated by and designed to test for coding errors, ensure accurate calculations, compare how differences between codes impact the results, and test the relative speed of different codes. The TRUST effort is focusing on the first three goals as testing the speed of codes is of much lower priority than ensuring that the codes are accurate, error free, and produce consistent results.

Ideally a full analytic solution would be adopted as the target solution for the TRUST benchmarks as this would allow for all benchmarking goals to be achieved. Unfortunately, no such analytic solutions exist for the dust radiative transfer equation for any geometries that are intrinsically 3D. Thus for 3D dust RT we are left with using a converged solution that has been validated by multiple codes, ideally with different solution techniques. This should ensure that the benchmark solution is not affected by coding errors and is likely to be correct. This paper presents the results for a geometrically simple, yet still 3D, benchmark. This first simple benchmark is an externally illuminated slab of dust and is presented specifically to test the components of dust RT at the basic level. Future benchmarks will test specific capabilities of codes in more complex geometries including shells, filaments, shadowed regions, and galaxy disks with spiral structures.

2. Setup

The overall geometry of this benchmark is a rectilinear slab that is externally illuminated by a single blackbody source. The values for the slab and point source are given in Table 1 and the geometry illustrated in Fig. 1. The luminosity of the source is set by integrating the source spectral energy distribution (SED) between 0.09 and 2100 μm. All the dust in the system is uniformly
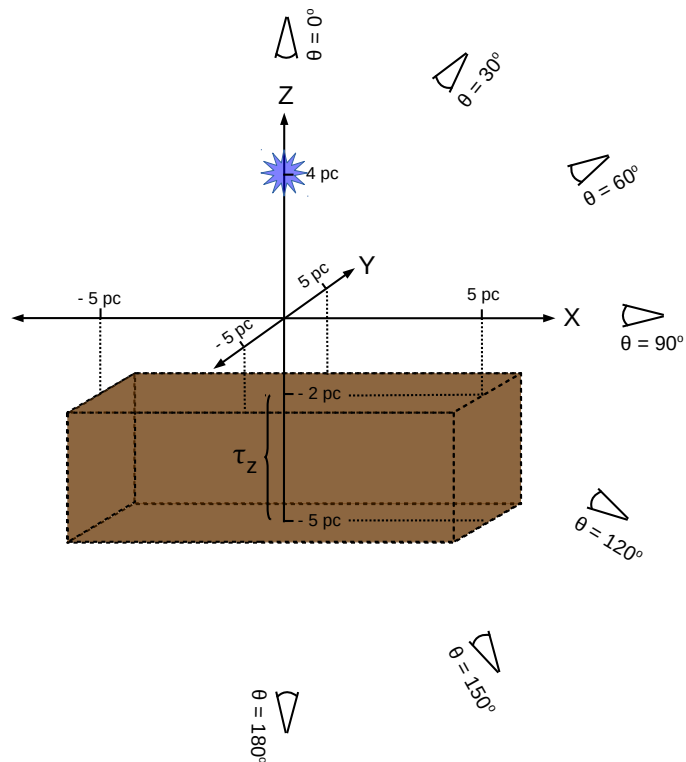
distributed through the slab with the rest of the model space being completely empty except for the blackbody source.

The dust grains have the BARE-GR-S properties from Zubko et al. (2004) and are discussed in detail by Camps et al. (2015). Given the computational complexity of including the full treatment of dust emission in the RT solution, this benchmark provides results including stochastic heating (sto, the full solution) as well as the equilibrium only heating (equ) and the single effective grain (eff) approximations. In the effective grain approximation, the grain properties are integrated over the grain size distribution and summed over the grain components to produce a single grain with effective properties. This effective grain is an extreme approximation (Steinacker et al. 2013) that has the benefit of allowing fast calculation of the dust emission spectrum from the radiation field. The equilibrium-only dust emission approximation assumes all grains are in equilibrium with the radiation field, even for those smaller grains that, physically, should be stochastically heated. The differences in the global SED from a model with the three different methods for calculating the emission from the dust are illustrated in Fig. 2. This figure clearly illustrates the importance of including the full solution to achieve accurate mid-IR and, to a lesser extent, far-IR results for cases like those in this benchmark. Data files that contain the full grain properties are available online[2] (Camps et al. 2015).

The density of the slab is set by the optical depth along the z-direction, defined as $\tau_z(1\,\mu m)$. The $\tau_z(1\,\mu m)$ values chosen here (Table 1) provide a full sampling of optical depths from optically thin to thick. For $\tau_z(1\,\mu m) = 0.01$, the model is optically thin at all wavelengths with a maximum of $\tau_z(0.09\,\mu m) \sim 0.18$. The next $\tau_z(1\,\mu m)$ value of 0.1 is optically thick in the UV with $\tau_z(0.09\,\mu m) \sim 1.8$. For $\tau_z(1\,\mu m) = 1$, the slab is optically thick to all UV and optical photons. Finally, the $\tau_z(1\,\mu m) = 10$ case is

---

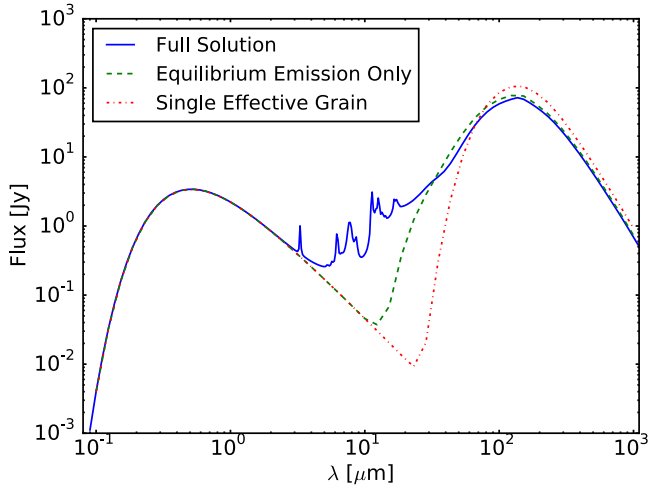**Fig. 2.** Global SEDs are shown with the three different dust emission choices. The example shown is for $\tau_z(1\,\mu m) = 1$ and $\theta = 90°$.

optically thick for all $\lambda \lesssim 4\,\mu m$ and very optically thick in the UV $\tau_z > 100$.

2.1. Requested outputs

Each RT code generates global SEDs and spatially resolved, multi-wavelength images. These outputs from each code are compared against all the other codes to estimate the fidelity of the RT solution for each configuration. We choose to use outputs for comparison as these quantities are what is generally compared with observations and the internal representation of quantities in RT codes (e.g., radiation field density) are often stored on quite different spatial grids. Previous benchmarks have also compared the internal dust temperature structure. We do not as only for the eff grain approximation is a single dust temperature defined. For the other two cases (equ and sto), there are a range of dust temperatures in each grid cell given the range of grain sizes and compositions. Global SEDs are computed for the total as well as decomposed into the different RT components (direct stellar, scattered stellar, direct dust emission, and scattered dust emission).

The SEDs are output in units of Jy and images are output in units of MJy/sr. SEDs and images are generated at seven viewing angles (0°, 30°, 60°, 90°, 120°, 150°, and 180°; see Fig. 1), probing the full range of scattering angles from back-scattering ($\theta = 0°$) to forward scattering ($\theta = 180°$). At lower optical depths ($\tau_z(1\,\mu m) = 0.01, 0.1, 1.0$), the resolution of the images is 300×300 pixels while at $\tau_z(1\,\mu m) = 10$, the image resolution is 600×600 pixels. The resolution is set by the need to resolve RT effects in the front surface of the dust slab in the infrared. In all cases, the physical size covered by the images is 15×15 pc to account for all possible rotations of the system. The two wavelength grids used are shown in Fig. 3 and can be obtained from the TRUST website[3]. The BASIC wavelength grid is used for the effective grain and equilibrium only emission approximations as the equilibrium grain temperatures only depend on the the total absorbed energy. The FULL wavelength grid is used for the full emission solution as the calculation of the stochastically heated grains requires a finer resolution sampling to calculate the temperature probability distribution. The FULL wavelength solution includes a very fine spacing in the mid-IR to resolve the aromatic/PAH emission features. Finally, we give
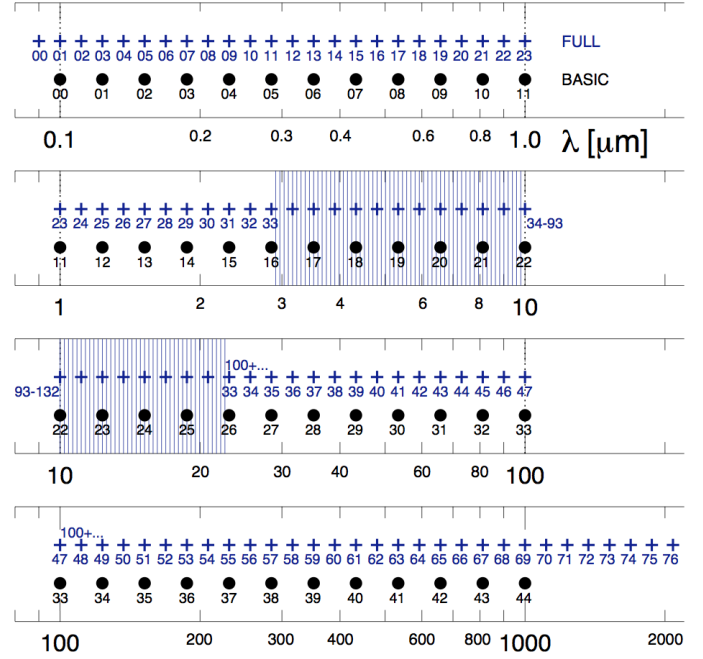
[3] http://ipag.osug.fr/RT13/RTTRUST/

Fig. 3. BASIC and FULL wavelength grids are shown. The BASIC grid is used for the models using the effective grain and equilibrium-only emission approximations. The FULL grid is used for the models computing the full dust emission including stochastically heated grains. Vertical blue lines between ~3–23 μm represent the dense sampling of wavelength points to resolve the PAH emission.

Table 2. Physical constants.

Constant	Description
$c = 2.99792458 \times 10^8$ m s^{-1}	speed of light
$h = 6.62606957 \times 10^{-34}$ m^2 kg s^{-1}	Planck constant
$k = 1.3806488 \times 10^{-23}$ m^2 kg s^{-2} K^{-1}	Boltzmann constant

the adopted values of relevant physical constants in Table 2 as their exact values will change the output SEDs and images.

2.2. Example outputs

Figure 4 gives an example of the global SED outputs. On the left, the total SED (Total) is shown along with the different components that comprise the total. Decomposing the total SED into components is important to test the differences in the different parts of the dust RT solution between models. The components include the stellar flux attenuated by any line-of-sight dust (Direct Stellar), the stellar flux scattered by the dust (Scattered Stellar), the thermal emission from the dust (Direct Dust Emission), and the scattered dust emission (Scattered Dust Emission). In addition, the stellar flux from the dust-free model (Transparent Stellar) is also shown as it is diagnostic of how each code treats the input stellar photons. The particular example shown illustrates the importance of the dust scattered stellar component in the ultraviolet and optical. The right panel shows the total SEDs covering the full $\tau_z(1\,\mu m)$ and θ range. The $\theta = 180°$ SEDs at shorter wavelengths clearly illustrate the impact of observing the star through the dust slab. The impact of dust self-absorption can be seen at $\tau_z(1\,\mu m) = 10$ most easily from the increasing depth of the silicate absorption feature at ~10 μm with increasing viewing angle θ.
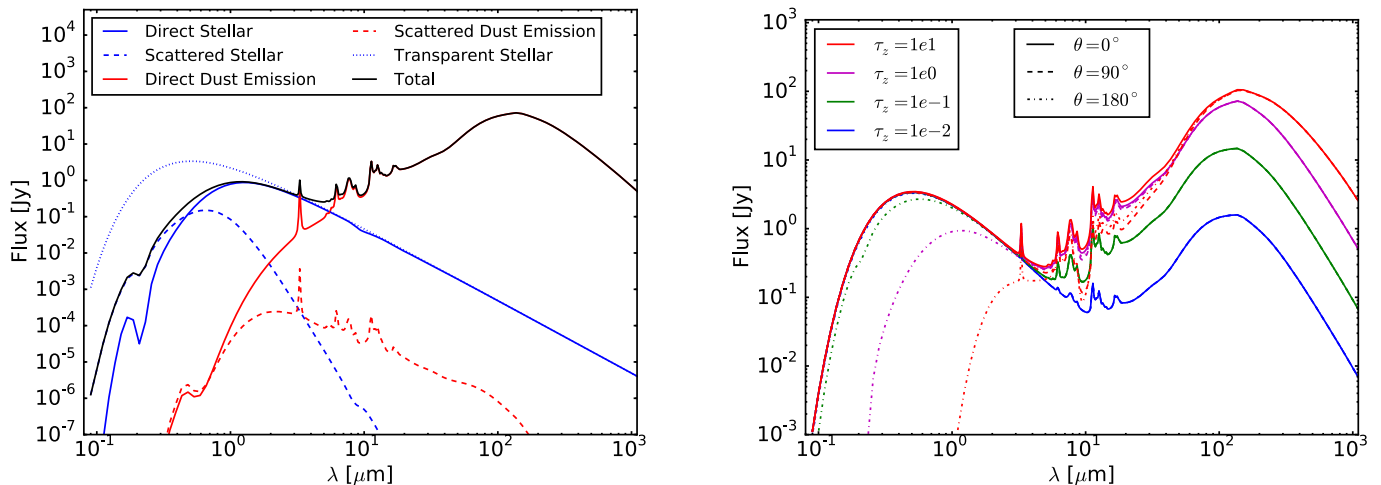
Fig. 4. Examples SEDs are shown from models run with the full dust emission solution. On the *left*, the total and components of the global SED are shown for the $\tau_z(1\ \mu m) = 1$ and $\theta = 150°$ case. On the *right*, the total SEDs are shown for all $\tau_z(1\ \mu m)$ values and θ values of $0°$, $90°$, and $180°$.
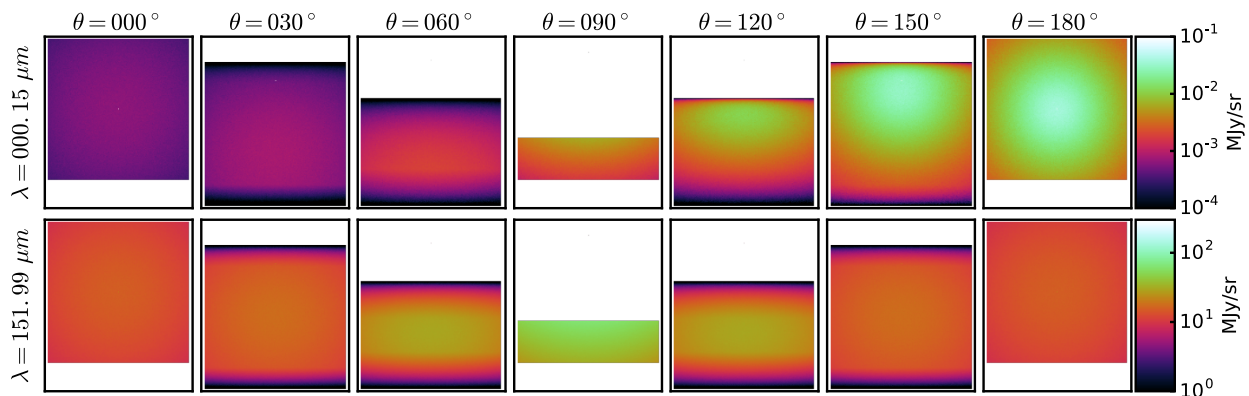

Fig. 5. Output images from the SKIRT model are shown for the $\tau_z(1\ \mu m) = 0.1$ case with the full dust emission solution for two representative wavelengths.

Examples of the output images are shown in Fig. 5 for the full range of viewing angles θ. The output SED at $\lambda = 0.15\ \mu m$ is dominated by scattering and the corresponding images illustrate the strongly forward scattering nature of dust grains with the brightness of the scattered light increasing from $\theta = 0°$ (back scattering) to $\theta = 180°$ (forward scattering). The $\lambda = 151.99\ \mu m$ images show that the infrared emission has a different dependence on viewing angle compared to the $\lambda = 0.15\ \mu m$ images, symmetrically peaking at $\theta = 90°$, reflecting the isotropic nature of the dust emission.

All the model outputs for all the codes are available online[4].

3. Models

There are seven 3D dust RT codes that participated in this benchmark. Six of them are based on Monte Carlo techniques (CRT, DIRTY, Hyperion, SKIRT, TRADING, and SOC) and one is based on Ray-Tracing (DART-Ray). Here we provide a very brief description of each code. The reader is encouraged to read the references provided for the details of how each code implements the solution to the RT problem.

3.1. CRT

CRT is a 3D Monte Carlo radiative transfer code (Juvela & Padoan 2003; Lunttila & Juvela 2012). It uses nested Cartesian

grids for representing the dust distribution. The program can use several methods for accelerating calculations, for example packet weighting (Juvela 2005), polychromatic photon packets (Jonsson 2006), and subgrid iteration (Lunttila & Juvela 2012). CRT includes a high-performance library for computing emission from arbitrary dust models, including stochastically heated grains (see Camps et al. 2015); however, it also allows using an external program for dust emission calculations (see, e.g., Ysard et al. 2011). The main application of CRT has been studies of molecular clouds and Galactic star formation. In particular, it has often been used for creating synthetic scattered light and dust thermal continuum observations of cloud models from magnetohydrodynamic simulations to help quantify the uncertainty in observationally derived cloud properties such as column density and core mass distribution (e.g., Juvela et al. 2006; Malinen et al. 2011). Other applications have included stability analysis of non-isothermal Bonnor-Ebert spheres (Sipilä et al. 2011) and galaxy mergers (Karl et al. 2013).

3.2. DART-Ray

DART-Ray is a purely ray-tracing 3D dust radiative transfer code. Its core algorithm has been presented by (Natale et al. 2014) and further developed by Natale et al. (in prep.). This algorithm reduces the amount of calculations that would be required in a brute-force ray-tracing approach by limiting the propagation of rays from each radiation source throughout the RT model

---

within the source influence volume. The latter is the volume around a radiation source within which the source contributes significantly to the radiation field energy density. This code utilises adaptive Cartesian grids to define the distributions of stars and dust and an optimization technique to set the angular density of rays cast from each source. The dust emission can be calculated both for dust in equilibrium with the radiation field and dust that is stochastically heated (Natale et al. 2015). The current version of DART-Ray does not include dust self-heating and dust emission scattering. In the context of this benchmark, the results are expected to be accurate in the infrared only for the cases where the dust emission is optically thin. We note that, in contrast with the other participating codes, the maximum number of scattering iterations is not set by the user in DART-Ray. Instead, scattering iterations are stopped when the remaining scattered radiation luminosity to be processed is less than 1% of the initial scattered radiation luminosity.

3.3. DIRTY

The DIRTY radiative transfer code is a 3D Monte Carlo radiative transfer code (Gordon et al. 2001; Misselt et al. 2001). It can handle arbitrary geometries efficiently using nested Cartesian grids. It includes a full dust grain model description allowing for arbitrary dust grain models to be used. The dust emission calculation includes the full solution including both equilibrium and non-equilibrium (stochastically heated) emission. This code has been used to study the general behavior of radiative transfer through clumpy dust (Witt & Gordon 1996), Milky Way reflection nebulae (Gordon et al. 1994; Calzetti et al. 1995; Lewis et al. 2009), regions of galaxies (Witt & Gordon 2000; Misselt et al. 2001), starburst galaxies locally (Gordon et al. 1997, 2000; Law et al. 2011) and at high redshift (Vijh et al. 2003), and disk galaxies (Pierini et al. 2004). For this benchmark, the spacing in the z direction was log and linear in the x and y directions. Motivated by this benchmark, the composite biasing technique (Baes et al. 2016) was added to better sample scattering at high optical depths and the dust emission from radiation fields with strong spatial variations.

3.4. Hyperion

Hyperion (Robitaille 2011) is an open-source[5] dust continuum 3D Monte Carlo radiative transfer code. It is designed to be modular, and can be used to compute radiative transfer for arbitrary 3D geometries and has been applied to, for example, analytical models of forming stars (Robitaille & Whitney 2014; Koepferl et al. 2015; Johnston et al. 2015), galaxy formation simulations (Narayanan et al. 2015), and large-scale Galactic emission (Robitaille et al. 2012). A number of grid geometries are supported, including Cartesian, spherical and cylindrical polar, nested Cartesian, octree, and Voronoi grids. Grid cells are never required to be regularly spaced, so for the models presented here, the cells in the z direction are spaced logarithmically (with the first cell below the surface of the slab located at -2.001). Multi-process parallelization is implemented using MPI and used for this benchmark. Hyperion supports computing the radiative transfer for one or more dust populations, but does not include full support for stochastic heating (instead, computing models with small grains and PAHs can be done using precomputed templates, as done in Robitaille et al. 2012). Since the original implementation of the code presented

in Robitaille (2011), forced first scattering using the Baes et al. (2016) algorithm has been added (and the results here assume $\xi = 0.2$). In addition, the process for monochromatic radiative transfer (described in Sect. 2.6.4 of Robitaille 2011) has now changed – when computing the contribution to the scattered light emission, photon packets are scattered and lose energy until their energy goes below a certain threshold (set to 10^{-120} for this paper), as opposed to randomly sampling whether to scatter or absorb at each interaction (which caused many scattering scenarios to be under-sampled and therefore required large numbers of photons in order to attain a good signal-to-noise in certain situations).

3.5. SKIRT

SKIRT is a public[6] multi-purpose 3D Monte Carlo dust radiative transfer code (Baes et al. 2011; Camps & Baes 2015) for simulating the effect of dust on radiation in astrophysical systems. It offers full treatment of absorption and multiple anisotropic scattering by the dust, computes the temperature distribution of the dust and the thermal dust re-emission self-consistently, and supports stochastic heating of dust grains (Camps et al. 2015). The code handles multiple dust mixtures and arbitrary 3D geometries for radiation sources and dust populations, including grid- or particle-based representations generated by hydrodynamical simulations. The dust density distribution is discretized using one of the built-in dust grids, including state-of-the art octree (Saftly et al. 2013), k-d tree (Saftly et al. 2014), and Voronoi (Camps et al. 2013) grids. The wide range of built-in components can be configured to construct complex models in a parameter file or through a user-friendly interface (Camps & Baes 2015; Baes & Camps 2015). SKIRT implements hybrid parallelization, allowing an arbitrary mix of multiple threads and processes possibly across multiple computing nodes (Verstocken et al., in prep.). While SKIRT is predominantly used to study dusty galaxies (e.g., Baes & Dejonghe 2002; Baes et al. 2010, 2011; De Looze et al. 2012; De Geyter et al. 2014; Saftly et al. 2015), it has also been applied to active galactic nuclei (Stalevski et al. 2012), molecular clouds (Hendrix et al. 2015), and stellar systems (Deschamps et al. 2015).

3.6. TRADING

TRADING (Bianchi 2008) is a 3D Monte Carlo radiative transfer code originally designed to study the effects of clumping in simulations of dust extinction and emission in spiral galaxies. Developed from an earlier regular-grid, thermal-emission-only code (Bianchi et al. 1996, 2000), TRADING includes an octree grid for the dust distribution, stochastic heating, and dust self-absorption. It neglects scattering of dust-emitted radiation. It has been used to model dust extinction (Bianchi 2007) and emission (Bianchi & Xilouris 2011; Holwerda et al. 2012) in edge-on galaxies, and to study the dust heating in the host galaxy of high-z Quasi-Stellar Objects (Schneider et al. 2015). A few adaptations were made to TRADING for this benchmark: while each photon packet wavelength is drawn from a (tabulated) probability distribution function (PDF), at odds with other RT codes, the weight of the photons was adjusted so that a similar number of packets falls in each bin of a logarithm-spaced grid (i.e., a weight $1/\lambda$ was used). Forced scattering was used for the first scattering event and along all paths with $\tau > 5$, applying the composite biasing scheme described by Baes et al. (2016) for

---

[5] http://www.hyperion-rt.org/

[6] http://www.skirt.ugent.be

Table 3. Code parameter values.

Name	Description	CRT	DART-ray	DIRTY	Hyperion	SKIRT	TRADING	SOC
N	# of photons/rays per wavelength	10^9	10^9–10^{10}	10^8	10^9	10^8	2×10^8	5×10^8
n_{xy}	# of bins in x or y	100	90–992	100	100	100	40–160	150
n_z	# of bins in z	150–300	29–31	100	200	100	144	150
m_{scat}	max # of scatterings	~60	8	500	~270	~250	~390	20
m_{iter}	max # of dust heating iterations	5	0	4	5	50	4	4

half the events along these paths. The models presented here were all run using an octree grid of $\approx 7.5 \times 10^5$ cells, having a higher resolution in the part of the slab facing the source.

3.7. SOC

SOC is a new Monte Carlo radiative transfer code that is parallelized using OpenCL libraries (Juvela et al., in prep.). The models can be defined on regular Cartesian, modified Cartesian, or octree grids. In this paper, calculations employ modified Cartesian grids that are defined by cell boundaries located at arbitrary positions along the three main axes. In dust-scattering calculations, SOC uses photon packages that each contain photons from four consecutive frequency bins. The differences between the optical depths and scattering functions at those frequencies are compensated by weighting (see Juvela 2005; Jonsson 2006). Calculations for dust emission proceed one frequency at a time. The dust temperature distributions and the dust emission (per cell) of stochastically heated grains are calculated with an external program that is common with CRT. For dust models assuming an equilibrium temperature, calculations are done within SOC itself.

4. Output comparisons

The comparison between the results from the different codes is based on measuring the average absolute deviation from the median result from all the codes. Explicitly, for each code (j) we calculate

$$\bar{\Delta}_j = \frac{1}{n} \sum_i |(x_{ij} - \mu_i)|, \quad (1)$$

where n is the number of wavelength/spatial points, x_i is the ith wavelength/spatial point for the global SED/image, and

$$\mu_i = \text{med}_j(x_{ij}). \quad (2)$$

We use the median for comparison as there is no analytic solution and use of a median is more robust to outliers. Given this metric, the deviation of one code from the reference being smaller than the deviation of other codes is not significant per se; it simply means that it is nearer the median. The goal of these quantitative comparisons is to determine the overall agreement between codes and to determine whether or not the differences between code results are random or systematic.

As certain parameters are particularly important for the precision of the results, as revealed by the convergence tests (Sect. 5), we give the values for such parameters for all the codes in Table 3. The number of rays/photons per wavelength (N) is only roughly comparable between different codes as, for example, the emission from the stellar source may be biased towards the slab or not. In addition, rays or photons may be split multiplying the impact of a single ray or photon. The value of m_{scat} is also only roughly comparable between codes as the scattering

may be done by forcing no scattering, forcing the first scattering, or forcing all scatterings. The maximum number of heating iterations m_{iter} controls whether or not dust self-heating is taken into account and how many iterations are allowed to achieve a specified convergence level. If $m_{iter} = 0$, then dust self-heating is not accounted for and the dust emission is only determined by the absorbed stellar photons. While these parameters are only roughly comparable between codes, they do provide a qualitative comparison between codes of the precision of different radiative transfer solution components.

While most of the codes include scattering of the dust emission, DART-ray and TRADING do not. This can affect the accuracy of the infrared images from these codes, but it is unlikely to affect the global SEDs as the total scattered dust emission is negligible compared to the total direct dust emission.

The adopted distance was not used fully in creating the images by all the different codes. Some codes took into account the projection effects due to the finite distance while others assumed an infinite distance resulting in no projection affects. The resulting differences are small in the images, except for the pixels sampling the edge of the slab and are noticeable for a viewing angle of $\theta = 90°$. These edge pixels were not used for the image comparisons to avoid injecting differences that are purely geometrical into the comparisons.

4.1. Example comparisons

The comparisons between the results of the different codes was done both with the global SEDs and images as shown in Figs. 6 and 7. The global SED comparisons were done both for the total SED as well as the different components (Sect. 2.2) of the RT solution. The image comparisons included both side-by-side display of the images as well as quantitative comparisons using two slices, one in the X and one in the Y direction. The slices were averaged in the direction perpendicular to the slice. The comparison between models was focused on a comparison of the behavior of these averaged slices. The global SED comparisons were diagnostic of issues with the different components of the solution (e.g., dust scattering and emission). The image comparisons were diagnostic of general issues in creating the images as well as cases seen only for a limited range of parameters (e.g., UV images at high optical depths). Quantitative analysis of the images focused on the Y slices as these were diagnostic of systematic issues at all θ values while being more robust to noise associated with the number of photons/rays.

The comparisons for all cases are given in the Slab section of the TRUST website[7].

4.2. Precision goal

In contrast with problems that have an analytic solution, for problems such as 3D dust radiative transfer that require a
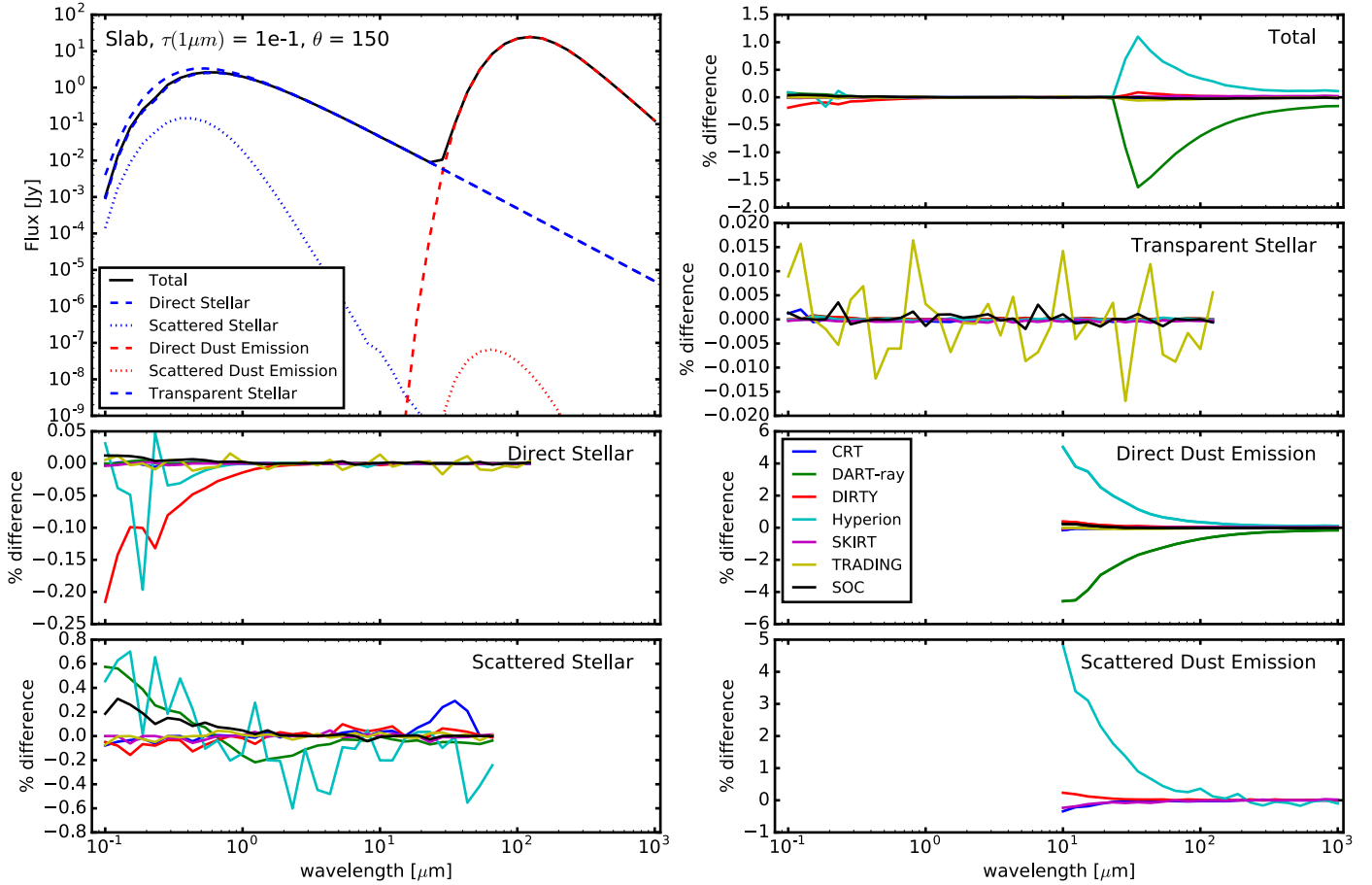
---

[7] http://ipag.osug.fr/RT13/RTTRUST/BM1.php

Fig. 6. An example of the model global SED outputs are shown for the $\tau_z(1\,\mu m) = 0.1$, $\theta = 150°$, and effective grain case. The *plot in the upper left corner* gives the median SED from all the models, both as a total and decomposed into components. The *other plots* give the percentage differences from the median for each of the components.

numerical approach, the solution will never converge fully to infinite precision. Thus, we need to define a robust metric to judge the convergence of the solutions presented here. One criteria could be when the convergence is well below the expected accuracy of the observations that the models are attempting to reproduce. Many observations are limited to accuracies of 1% or larger based on uncertainties in absolute flux or surface brightness calibration (e.g., Bohlin et al. 2011; Balog et al. 2014). Another approach is to use the differences seen in previous 2D dust RT benchmark as an upper limit. The differences between the global SEDs in the Pascucci et al. (2004) 2D disk geometry benchmark for all optical depths ($\tau(V) = 0.1-100$) was <10% and <3% for the the lowest optical depth. The Pinte et al. (2009) circumstellar disk geometry benchmark for high optical depths ($\tau(I) \approx 10^3 - 10^6$) found differences between codes on the order of 10%. Finally, available computational power imposes a limit – it should be possible to carry out the computations in a reasonable time. This allows for more codes to participate in this benchmark and makes it reasonable for new codes to use these benchmarks to test their accuracy. Combining all these points, we adopt a precision goal of 1% on the global SED at lower optical depths and a more relaxed 10% for higher optical depths. For the *Y* slice image-based comparisons, we have chosen a 10% precision goal.

4.3. Resulting precisions

The percentage differences between codes for the global SED components are summarized in Table 4 and for the *Y* slices at

Table 4. SED average deviations.

Component	$\tau_z(1\,\mu m) \leq 1$	$\tau_z(1\,\mu m) = 10$
Direct stellar	0.3%	1.7%
Scattered stellar	0.9%	3–58%
eff		
Direct dust emission	0.7%	4.0%
Scattered dust emission	0.7%	3.7%
equ		
Direct dust emission	0.3%	3.1%
Scattered dust emission	1.3%	3.3%
sto		
Direct dust emission	2.9%	3.2%
Scattered dust emission	2.8%	2.8%

specific wavelengths in Table 5. In general, we find the results to be within the goal precisions with the notable exceptions of the stellar scattered and dust emission components for $\tau_z(1\,\mu m) = 10$. The details of these comparisons are discussed below.

4.3.1. Stellar radiative transfer

While the properties of the calculated dust emission are sensitive to whether an approximation is used (eff or equ) or the full solution is calculated (sto) – see Fig. 2 – the radiative transfer problem itself is not. In fact, the radiative transfer of photons through dust is mathematically equivalent regardless of whether
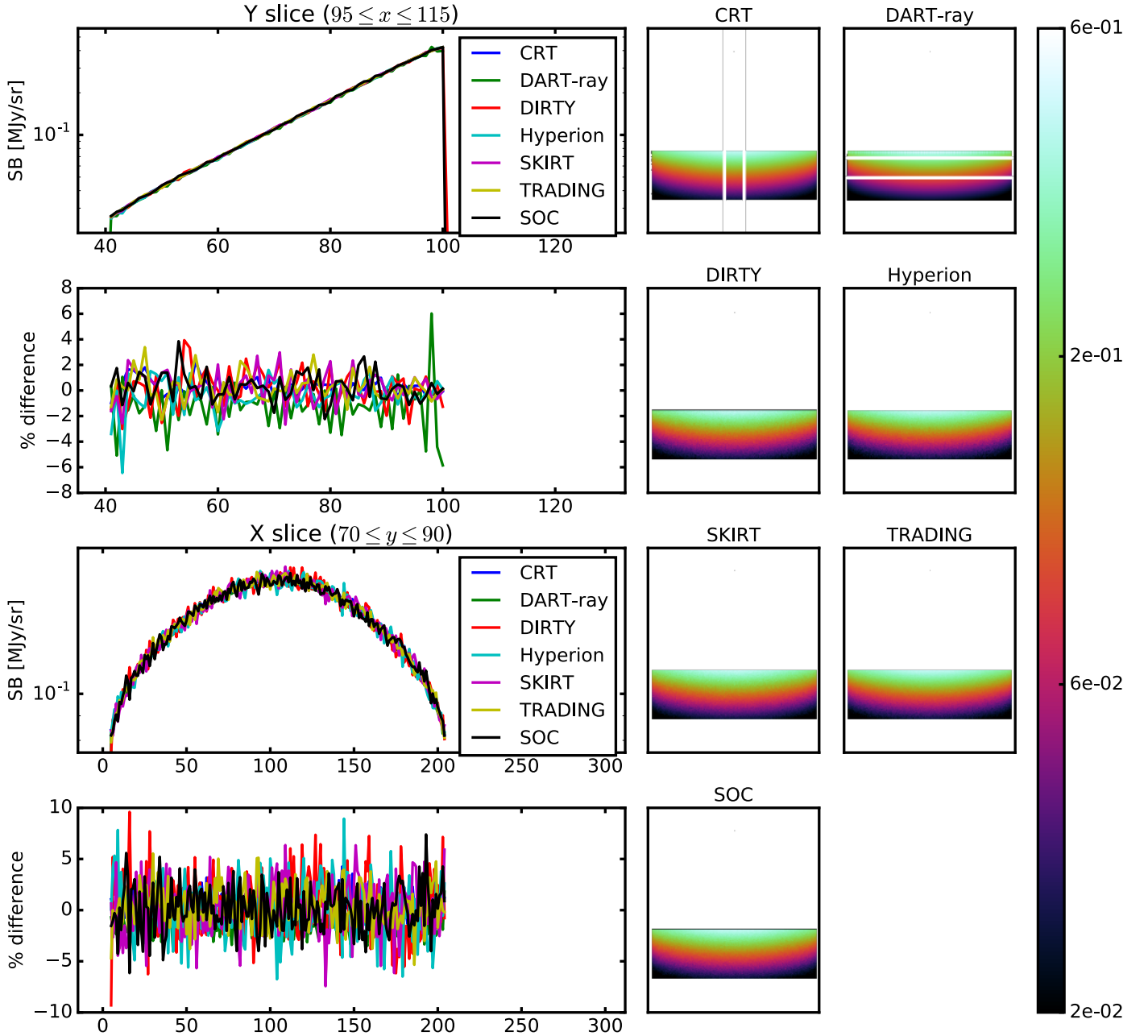
Fig. 7. An example of the model image outputs is shown for the $\tau_z(1\ \mu\text{m}) = 1$ and effective grain case. In addition to the total images for each model, *Y* and *X* slices are shown along with the differences for each model from the median slice. The *X* and *Y* slices refer to the output image dimensions, not the axes in the 3D model space. The locations of each slice are shown over-plotted on the 1st (*Y*-slice) and 2nd (*X*-slice) model images where the slice is computed as an average over the slice width.

the photon interaction is computed separately for each size and composition in a distribution, or computed for an effective grain generated by integrating the grain properties over the size distributions and chemical compositions (Steinacker et al. 2013). Hence, comparisons of the direct and scattered stellar light provide a comparison of each code's treatment of the radiative transfer problem, free from the additional computational complications arising from computing the dust emission.

The direct and scattered stellar comparisons for the global SEDs are shown in Fig. 8 and summarized in Table 4. The direct stellar component shows the largest differences at high θ values

where the slab occults the illuminating star. They are largest at $\theta = 150°$, grow with increasing $\tau_z(1\ \mu\text{m})$, and are just above the goal of 1% at all optical depths. The $\theta = 150°$ case provides the maximum optical depth from the star to the observer. The differences for $\theta < 150°$ are due to different numerical representations of the intrinsic SED between codes and are $\ll 0.1\%$. For the scattered stellar component, the difference is $<1\%$ for $\tau_z(1\ \mu\text{m}) \leq 1$ and 3% for $\tau_z(1\ \mu\text{m}) = 10$ with the exception of $\theta = 180°$ where the difference is 50%. Larger differences are not unexpected at $\theta = 180$ as there are no paths for scattered photons that do not require them to penetrate the entire slab, making the results very

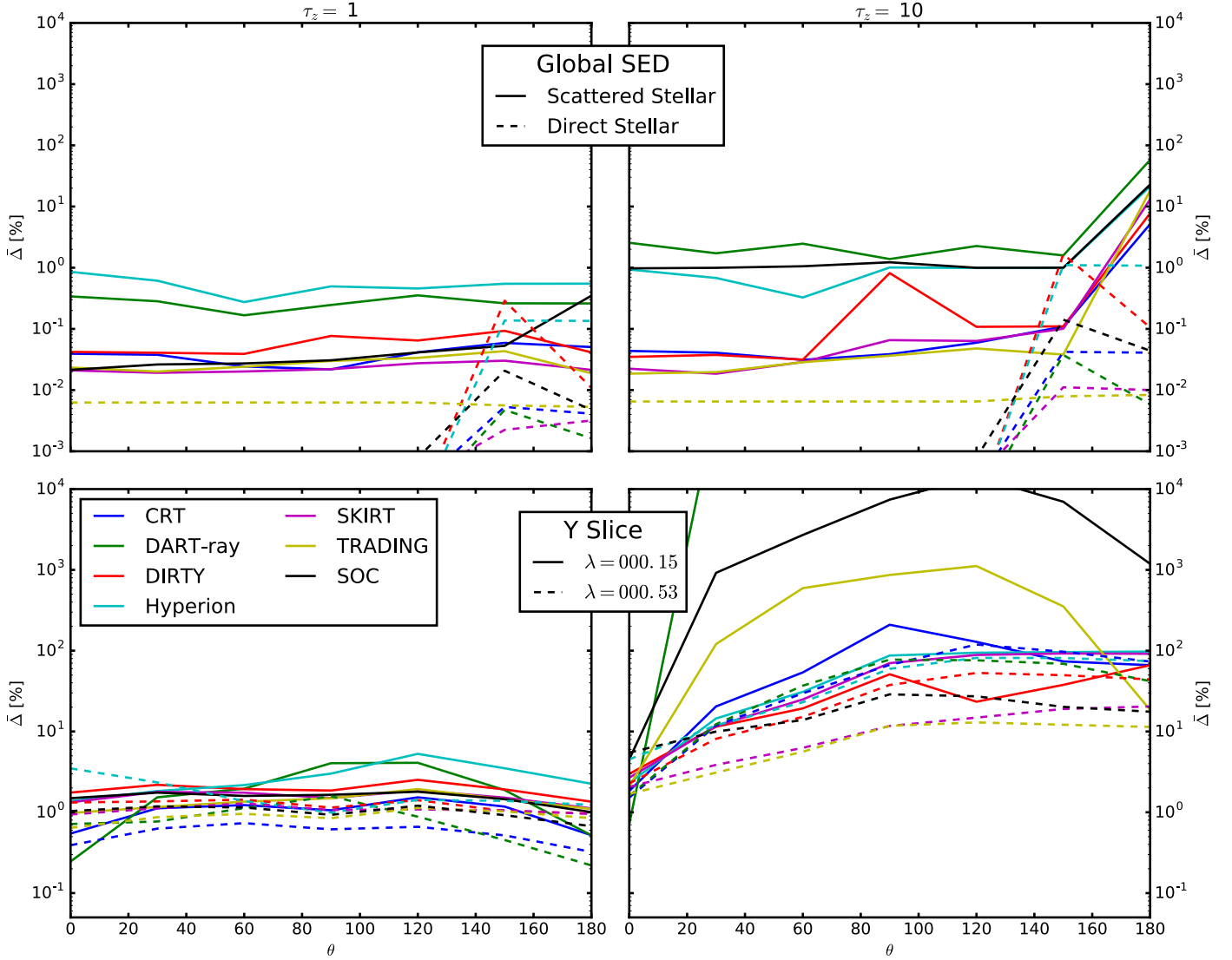**Fig. 8.** Average deviation of the global SEDs from the median results are shown versus θ for the direct and scattered stellar photons in the *top row*. The *Y* slice average deviation from the median results are shown versus θ for two UV/optical wavelengths probing the scattered stellar photons in the *bottom row*. The results for $\tau_z(1\ \mu m) = 0.1$ and 0.01 are similar to those for $\tau_z(1\ \mu m) = 1.0$.

Table 5. Y Slice average deviations.

$\lambda\ (\mu m)$	$\tau_z(1\ \mu m) \leq 1$	$\tau_z(1\ \mu m) = 10$
0.15	5.3%	$\gg 1000\%$
0.53	3.5%	119%
eff		
35.11	8.3%	3.9%
151.99	3.4%	2.1%
equ		
35.11	3.8%	5.0%
151.99	4.7%	0.9%
sto		
8.11	10.7%	8.7%
23.10	6.7%	19.9%
151.99	6.5%	4.1%

sensitive to the number of photon packets run for a given model. Thus, the $\tau_z(1\ \mu m) = 10$ and $\theta = 180°$ is a very sensitive test of the dust scattering at high optical depths.

The comparisons for the *Y* slices of the images at two representative wavelengths (one in the UV and one in the optical) are shown in the bottom row of Fig. 8 and summarized in Table 5. For $\tau_z(1\ \mu m) \leq 1$, the precisions are well within the goal of 10%. At $\tau_z(1\ \mu m) = 10$, the discrepancies are much, much larger than this goal. The plot shows a bifurcated behavior for $\theta > 0°$ that is the signature of very large variations between the median and all of the code results. The results that are below the median by a large value give $\bar{\Delta} \sim 100\%$ as they are effectively zero. The results that are above the median by a large value give $\bar{\Delta} \gg 10^4\%$. For $\theta = 0°$, all the codes have $\bar{\Delta} < 10\%$ as the scattered stellar images are dominated by back scattering off the surface of the slab. Physically, one would expect that, even for strongly forward-scattering grains at very high optical depth, there would be very little scattered stellar light in the models for viewing angles near $\theta = 180°$. Indeed, that is what is observed for this benchmark. However, given that the amount of scattered stellar light is very small, small differences in how different codes handle the scattering (and the RT problem in general) can lead to very large relative discrepancies, as observed. These results illustrate that 3D dust RT at high optical depth is
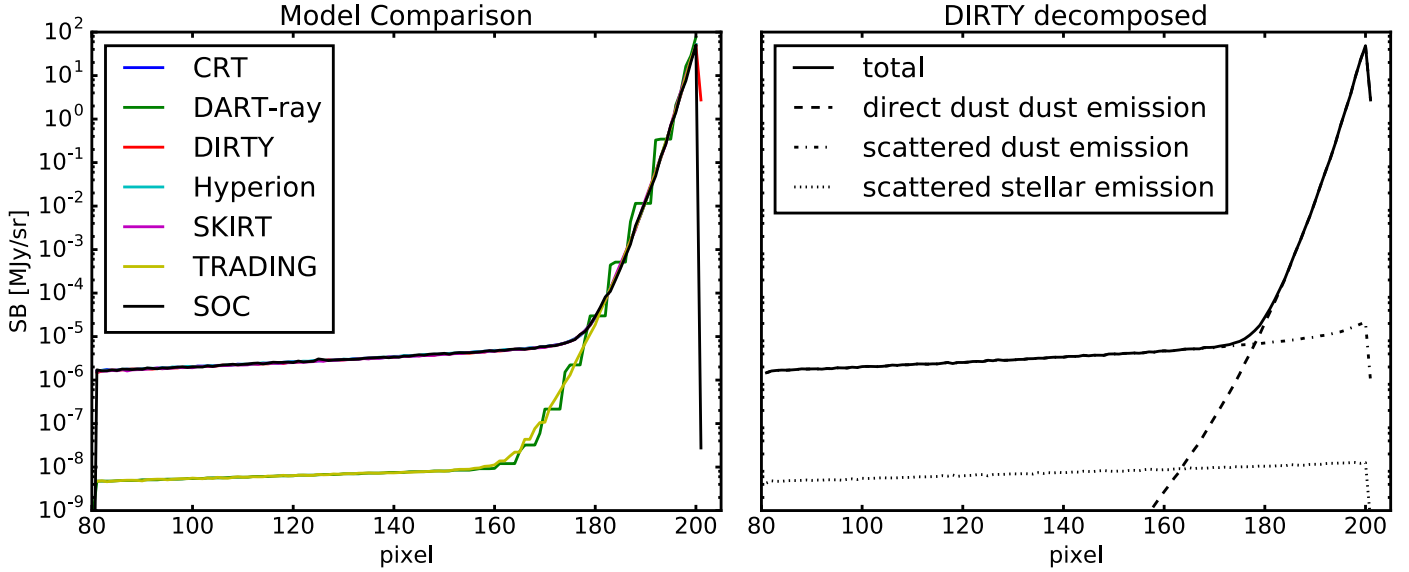
Fig. 9. Contribution the scattered dust emission at $\lambda = 35.11$ μm makes to the Y slice for $\tau_z(1\,\mu m) = 10$, $\theta = 90°$ is shown. The comparison between all the code results is shown on the *left*, where the DART-ray and TRADING results are significantly below the results from the other five codes. On the *right*, the results from DIRTY are shown decomposed into the three contributing components. For the back three quarters of the slab, the scattered dust emission dominates over the direct dust emission and scattered stellar emission. The differences between DART-ray and TRADING and the rest of the codes is due to not calculating the scattered dust emission component.

still a challenging numerical problem for RT codes and can provide a sensitive probe of the efficacy of the numerical solution implemented.

4.3.2. Dust emission

The dust emission changes significantly depending on the assumption used as shown in Fig. 2. The results for each of the dust emission approximations (eff = effective grain, equ = equilibrium only, and sto = full solution including stochastically heated grains) are given. Almost all of the codes provide results for all three dust emission cases. The exceptions are Hyperion, which only provided results for eff case results, and SOC, which does not provide the equ case results. In both cases, the codes could compute the missing cases, but time limitations meant they were not computed.

DART-ray and TRADING do not compute the dust-emission scattered component. The importance of the dust-scattered emission is illustrated in Fig. 9. In addition, DART-ray does not allow for the heating of dust due to its own emission. The importance of dust self-heating is discussed in Sect. 5.4. The lack of the full dust emission radiative transfer calculation in these two codes means that it is expected that their results will be less accurate for high optical depths. High optical depths are where the dust-emission scattering and self-heating are particularly important. For these reasons and to be conservative, we do not include DART-ray in determining the precisions of the solution for any optical depth for the global SED dust-emission components. In addition, we do not include DART-ray or TRADING for the $\tau_z(1\,\mu m) = 10$ results for determining the precision of the dust emission Y slices. We do include all the results in the figures, with those results not used for precision calculations shown as faint lines. For longer IR wavelengths the results for TRADING do not seem to be significantly affected by not including the dust-emission scattering calculation, but for the purposes of establishing the precision of the benchmark, we conservatively do not include them.

The direct and scattered dust-emission comparisons for the global SED components are shown in Figs. 10–12 for the eff, equ, and sto cases, respectively, and summarized in Table 4. The comparisons for the Y slices of the images at two representative wavelengths (one in the mid-IR and one in the far-IR) are shown in the bottom row of the same figures and summarized in Table 5. For the eff case, the precisions achieved are at or better than the goals. For the equ case, the precisions achieved meet the goals, except for the global- dust-emission scattered component that has a precision of 2%. For the sto case, the precision was 3% for the global dust-emission components, well above the goal of 1%. The Y slice precisions were better than the goal of 10%, except for the $\lambda = 23.10$ where the differences were at the 20% level.

While the goal precisions were achieved for the eff case, it is worth noting that the differences between the Hyperion results and most of the codes are likely caused by differences in the sampling of the photons' wavelengths. Specifically, most of the codes sample the photon wavelengths directly from the wavelength grid while Hyperion samples the photon wavelengths from a continuous spectrum and the wavelength grid is only used for output quantities. With a higher-resolution wavelength grid, it is likely that most of the codes would cluster around the Hyperion results.

5. Convergence tests

In the absence of an analytic solution, another method for building confidence in the numerical solution is to perform convergence tests. Such tests also provide insight into the effects different limits place on the solution (e.g., the importance of scattered photons or dust self-heating). These tests involve changing numerical tolerances and quantifying how the solution changes. Convergence testing is most often done based on an experience-based understanding of the relative importance of the model parameters in influencing the solution.

We have performed a number of quantitative convergence tests for this benchmark using the DIRTY code. Similar results
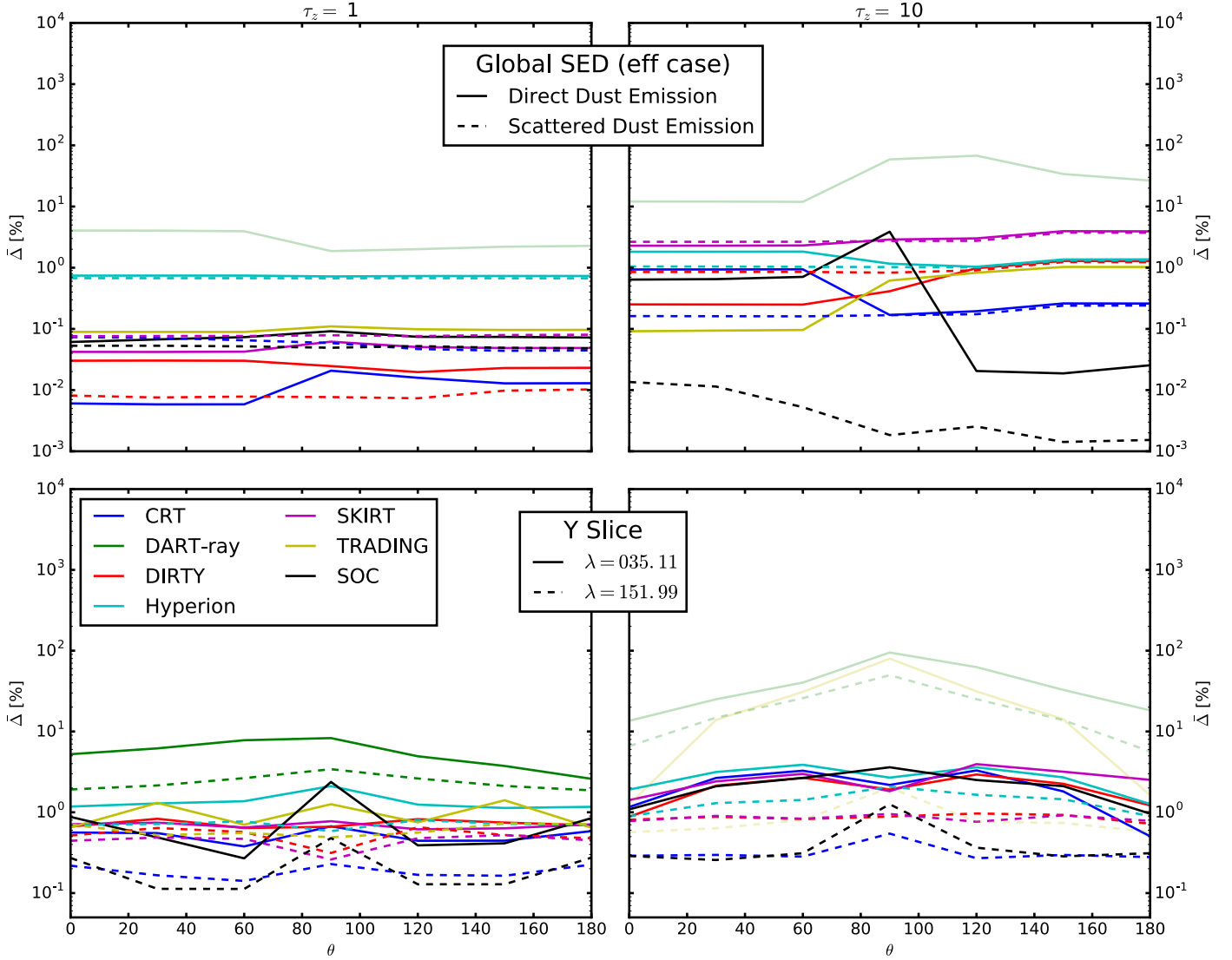
Fig. 10. Average deviation from the median results are shown versus θ for the direct and scattered dust emission for the effective grain approximation in the *top row*. The Y slice average deviation from the median results are shown versus θ for two IR wavelengths probing the dust emission in the *bottom row*. The results for $\tau_z(1 \, \mu m) = 0.1$ and 0.01 are similar to those for $\tau_z(1 \, \mu m) = 1.0$. Models not used in the precision calculation are shown as faint lines.

Table 6. DIRTY base parameter values.

Name	Values	Description
N	10^8	# of photons per wavelength
n_{xy}	100	# of bins in x or y
n_z	100	# of bins in z
m_{scat}	500	max # of scatterings
n_{iter}	4	# of dust heating iterations
ξ_{scat}	0.5	scattering composite bias
ξ_{emis}	0.5	emission composite bias

would be expected for the other Monte Carlo codes and, to a lesser extent, with Ray-Tracing codes. For the convergence tests, the parameters not being tested were set to values given in Table 6. We have not exhaustively searched the possible parameter space in our convergence tests, due to the significant computational resources necessary for each set of parameter tests. Instead, we have fixed all the parameters, except the one being

varied, to values that are expected to provide reasonable precision based on previous runs.

We have performed the convergence tests for $\tau_z(1 \, \mu m) =$ 1 and 10. Practically, we found that testing the precision for $\tau_z(1 \, \mu m) = 1$ could be done in a reasonable amount of computer time and the results for lower optical depths will have comparable or better precision. The $\tau_z(1 \, \mu m) = 10$ was challenging for all codes and the precision remains limited by computer time (i.e., <1 month or so single threaded being reasonable within the scope of this study).

5.1. Number of photons/rays

The number of photons or rays that are computed at each wavelength (N) is clearly a parameter that will strongly influence the precision of the model results. This model parameter controls the precision of the scattered light calculation and of the dust emission from each grid cell. A significant portion of the computations scale directly with N.
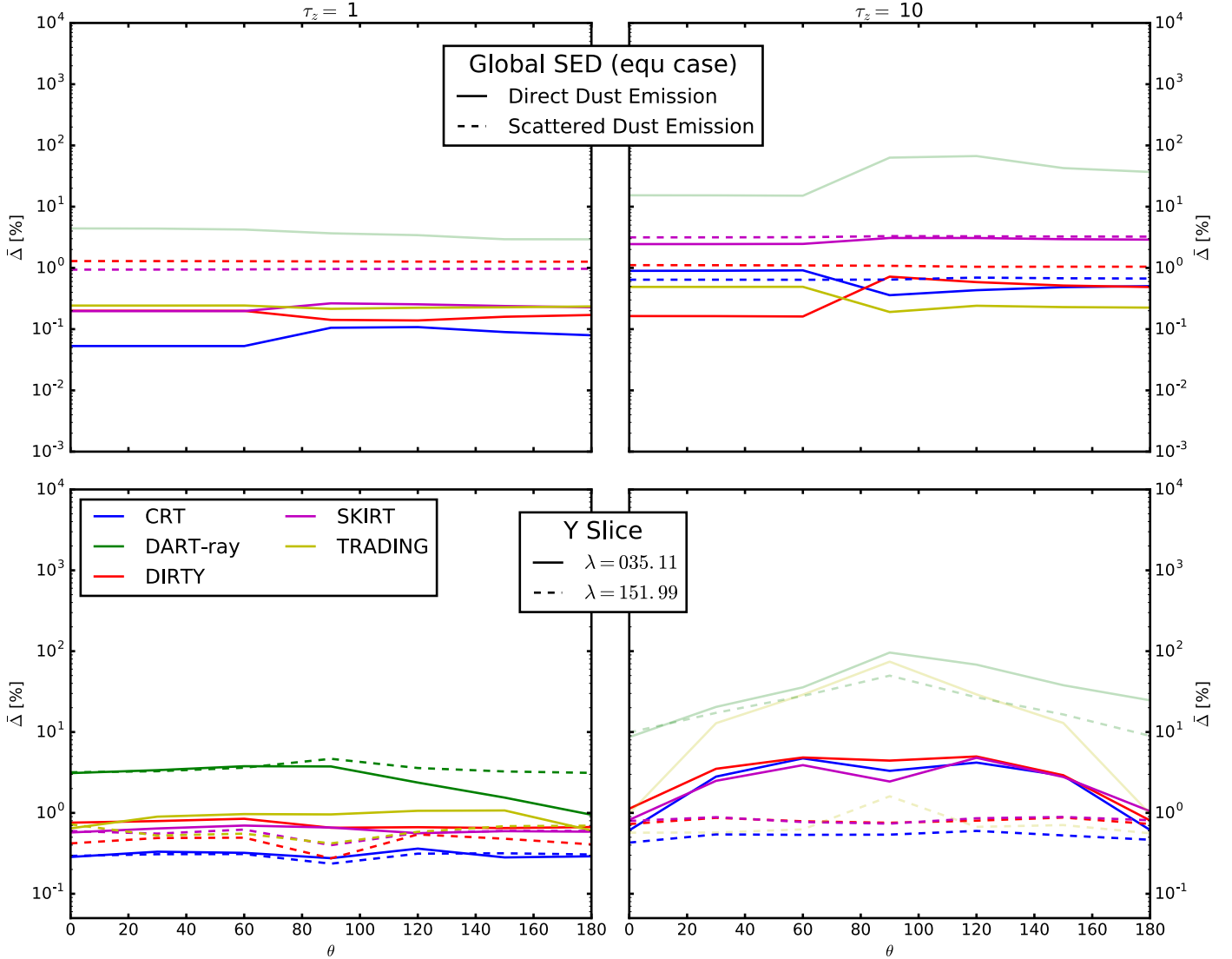
Fig. 11. Average deviation from the median results are shown versus θ for the direct and scattered dust emission for the equilibrium only grain approximation in the *top row*. The Y slice average deviation from the median results are shown versus θ for two IR wavelengths probing the dust emission in the *bottom row*. The results for $\tau_z(1\,\mu m) = 0.1$ and 0.01 are similar to those for $\tau_z(1\,\mu m) = 1.0$. Models not used in the precision calculation are shown as faint lines.

The N convergence tests were computed for three angles to illustrate the impact of N on backscattered photons (0°), penetration depth into the slab (90°), and penetration through the entire slab (180°). Figure 13 plots the average deviation as a function of N for the total global SED and the Y slices. The average deviation is computed compared to the model run with the largest N (e.g., 10^8). For $\tau_z(1\,\mu m) = 1$, convergence of the global SED and component SEDs (not shown) to 1% is achieved with $N = 10^6$ and $N = 10^7$, respectively. For $\tau_z(1\,\mu m) = 10$, a similar behavior is seen except for the scattered stellar component where convergence is not seen and the average deviation remains high (~20–80%) for all values of N tested. The convergence for the Y slices is more complicated. Convergence to 10% is reached by $N = 2 \times 10^7$ for small angles (e.g., 0°) at all wavelengths. For $\tau_z(1\,\mu m) = 1$, the goal convergence is seen for the IR wavelengths around 10^6 and for the UV/optical wavelengths at 10^7. For $\tau_z(1\,\mu m) = 10$, the goal convergence is again seen for the IR wavelengths around 10^6, but the convergence for the UV/optical wavelengths is well beyond the number of photons tested with an extrapolated prediction of convergence around $N > 10^9$ photons.

From these calculations, we can see that $N = 10^8$ will provide good precision for all but the UV/optical wavelengths for the $\tau_z(1\,\mu m) = 10$ case.

5.2. Spatial grid

Another obvious model configuration parameter to test for convergence is the number of spatial bins that describe the slab. The computation time required for a model scales with the number of bins through the need to compute the dust emission from each bin as well as computing the radiative transfer through all the bins. The finer the slab is divided, the more exact the solution becomes, but also the more time required for the computations. It is useful to note that using the Monte Carlo solution technique with a uniform dust density results in the scattering component of the solution being independent of the number of spatial bins used. This is because the scattering location is computed exactly for each photon packet independent of the grid.
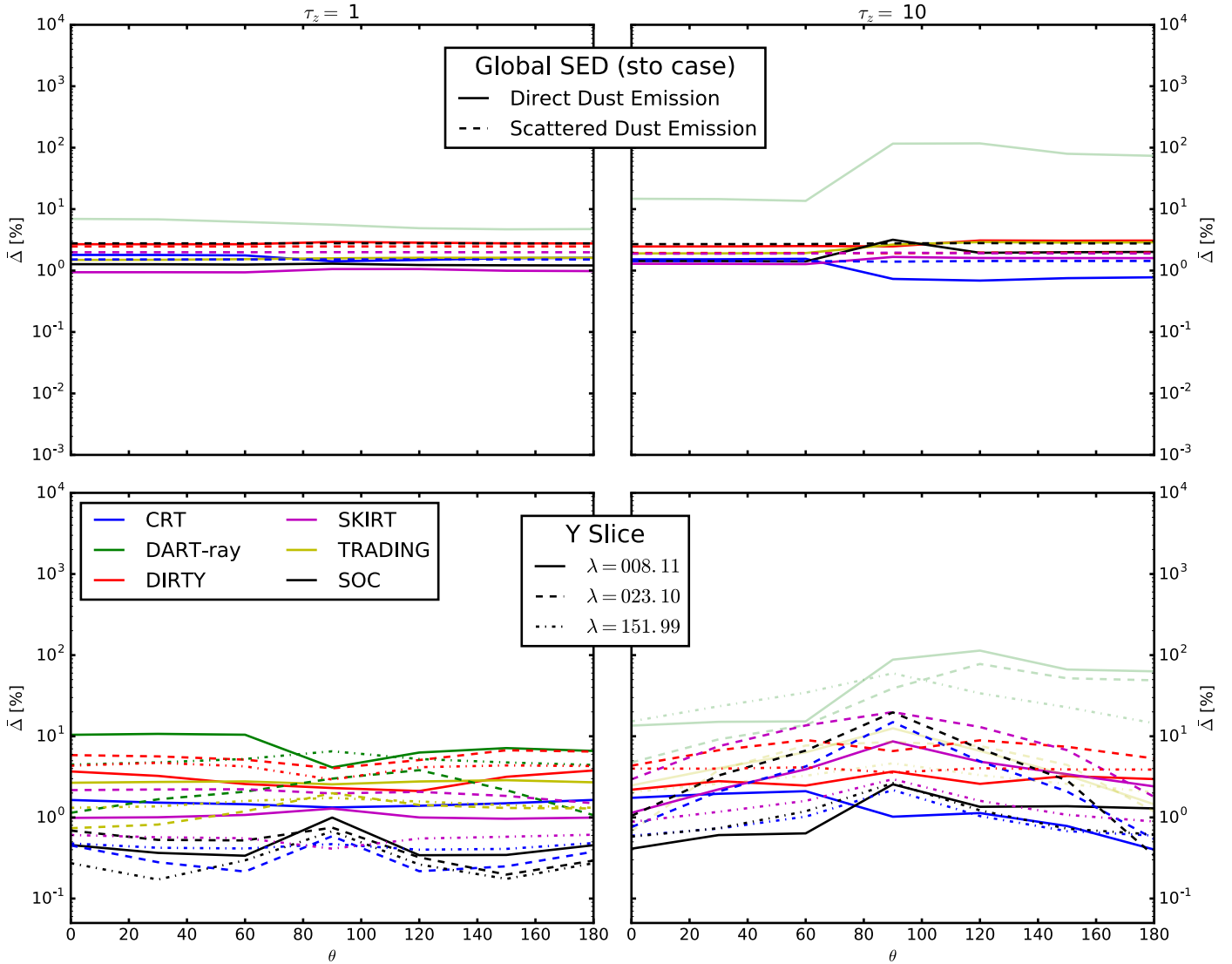
Fig. 12. Average deviation from the median results are shown versus θ for the direct and scattered dust emission for the full grain solution (equilibrium and non-equilibrium dust emission) in the *top row*. The Y slice average deviation from the median results are shown versus θ for three IR wavelengths probing the dust emission in the *bottom row*. The results for $\tau_z(1\,\mu m) = 0.1$ and 0.01 are similar to those for $\tau_z(1\,\mu m) = 1.0$. Models not used in the precision calculation are shown as faint lines.

The slab geometry naturally lends itself to simple division into bins in the x, y, and z dimensions. The convergence tests were done using linear x and y spacing and logarithmic z spacing starting from the slab front facing nearest the star. We found that logarithmic bin spacing in z provided equivalent results to linear z bin spacing but required fewer bins for the same precision.

The convergence tests for the number of x and y bins ($n_x = n_y \equiv n_{xy}$) were computed for the $\theta = 0°$ and $180°$ cases as these two viewing angles are the most sensitive to n_{xy}. Figure 14 plots the average deviation as a function of n_{xy} for the total global SEDs and Y slices. The average deviation is computed compared to the model run with the largest n_{xy}. Convergence of the global SEDs to 1% is achieved with $n_{xy} = 2$ for the total and all the components. Similarly, 10% convergence for the image slices is achieved by $n_{xy} = 10$.

The n_z convergence tests were computed for $\theta = 90°$ as this viewing angle is the most sensitive to n_z. Figure 15 plots the average deviation as a function of n_z for the total global SEDs and Y slices where average deviation is computed compared to the model run with the largest n_z. For $\tau_z(1\,\mu m) = 1$, convergence

of the global SEDs to 1% is achieved with $n_z \sim 4$. Convergence to 1% for the components of the global SEDs is achieved for $\tau_z(1\,\mu m) = 10$ at $n_z \sim 30$. The image slice convergence to 10% is achieved at $n_z \sim 20$ for $\tau_z(1\,\mu m) = 1$ and $n_z \sim 35$ for $\tau_z(1\,\mu m) = 10$.

5.3. Number of scatterings

Calculation of the scattered photons is one of the challenging parts of the radiative transfer solution, especially in light of the importance of multiple scattered photons at higher optical depths. In general, RT codes set a "maximum number of scatterings" to avoid photon packets getting "stuck", especially in very high optical depth environments. Of course, if that maximum value is set below the typical number of scatterings a photon might undergo, it can lead to erroneous results from the simulation for the scattered signal. To quantify the importance of multiple scattering, we carried out convergence tests at θ values of $0°$, $90°$, and $180°$. For $\tau_z(1\,\mu m) \leq 1$, the number of scatterings needed to achieve the goal precisions is on the order of 5.
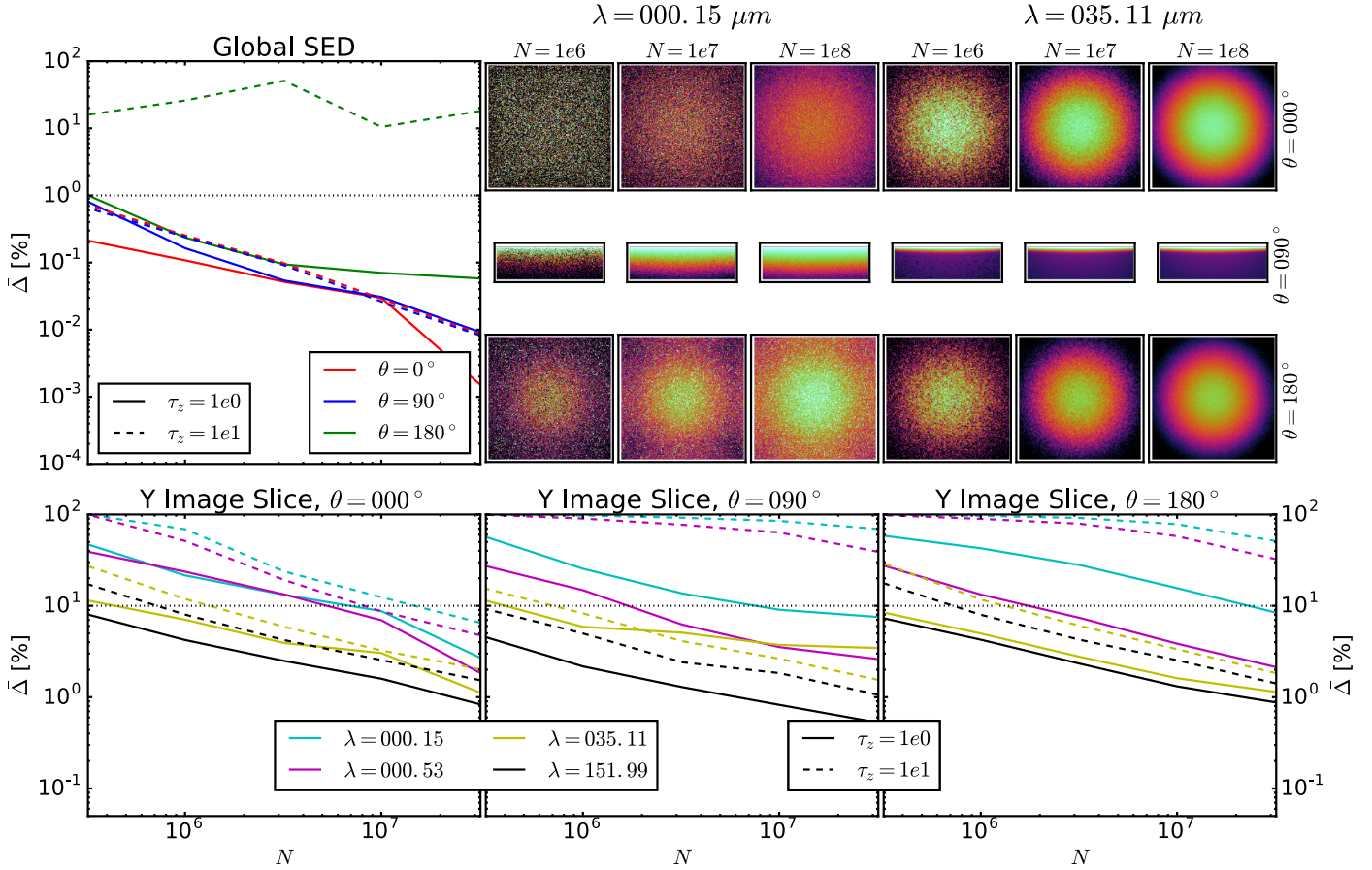
Fig. 13. Average deviations ($\bar{\Delta}$) versus the number of photons or rays that are computed at each wavelength (N) are shown for the total global SEDs and Y image slices. The images are for the most challenging case of $\tau_z(1\ \mu m) = 10$ and illustrate the qualitative impact of increasing N. Plots and images are shown for $\theta = 0°$, $90°$, and $180°$. The dotted horizontal lines give the 1% (global SED) and 10% (Y slices) levels. The images are plotted with the same log scaling for each wavelength and angle combination. The results for $N = 10^8$ are not shown as $\bar{\Delta}$ is computed relative to this case where $\bar{\Delta} = 0$ by definition.

For $\tau_z(1\ \mu m) = 10$, approximately 20 scatterings are needed for the goal precisions with this driven mainly by the convergence at the shortest wavelengths. We illustrate the importance of multiple scattering for the $\tau_z(1\ \mu m) = 1$ and 10 cases in Fig. 16. This figure gives the Y slices and percentage deviations from the $m_{scat} = 500$ case for $\lambda = 0.15$, 35.11, and 151.99 μm. We only show the 90° case as it clearly illustrates the changing importance of multiple scattering as a function of penetration depth in the slab. Multiple scattering is important for both the direct scattering at $\lambda = 0.15\ \mu m$ and the dust emission at both IR wavelengths. The dependence of the IR wavelengths on the number of scatterings clearly shows the importance of dust heating due to scattered photons. As expected, the dependence on multiple scattering is largest at higher optical depths.

5.4. Self-heating iterations

The thermalized radiation emitted by dust that is heated by the primary radiation source can in turn be absorbed by other dust grains in the model space, leading to dust self-heating. This self-heating increases in importance as the optical depth increases, since, depending on the location, the dominant radiation source may be the re-emitted dust emission. Most RT codes account for dust self-heating by iterating between the dust emission and the dust absorption and scattering, stopping when a preset energy convergence is achieved. We carried out convergence tests to

quantify the importance of dust self-heating. For $\tau_z(1\ \mu m) \leq 1$, both the global SED and Y-slice comparisons show that no iterations are necessary to meet our precision goals; the effect of dust self-heating is small enough that neglecting it does not change the precision of the resultant model. This is not the case for $\tau_z(1\ \mu m) = 10$ where neglecting dust self-heating results in global SEDs in error larger than the goal precisions for all cases and the Y slices for the $\theta = 90°$ case. The $\theta = 90°$ case clearly shows the impact of dust self-heating and we illustrate this in Fig. 17. This figure gives the Y slices for a range of dust self-heating iterations starting with no self-heating ($m_{iter} = 0$). The impact of dust self-heating is to raise the emission in the front of the slab at shorter wavelengths (e.g., $\lambda = 35.11\ \mu m$) and, more dramatically, over most of the slab at longer wavelengths (e.g., $\lambda = 151.99$). Fortunately, only a single dust self-heating iteration is needed, with additional self-heating iterations providing only small gains.

5.5. Scattering at high optical depths

Radiative transfer at high optical depths through dust is challenging for most numerical solution techniques. There are approximations possible such as the diffusion approximation (Kuiper et al. 2010), but such approximations impose real limitations on the accuracy of the resulting calculations. Motivated by the work on this benchmark, a new technique based
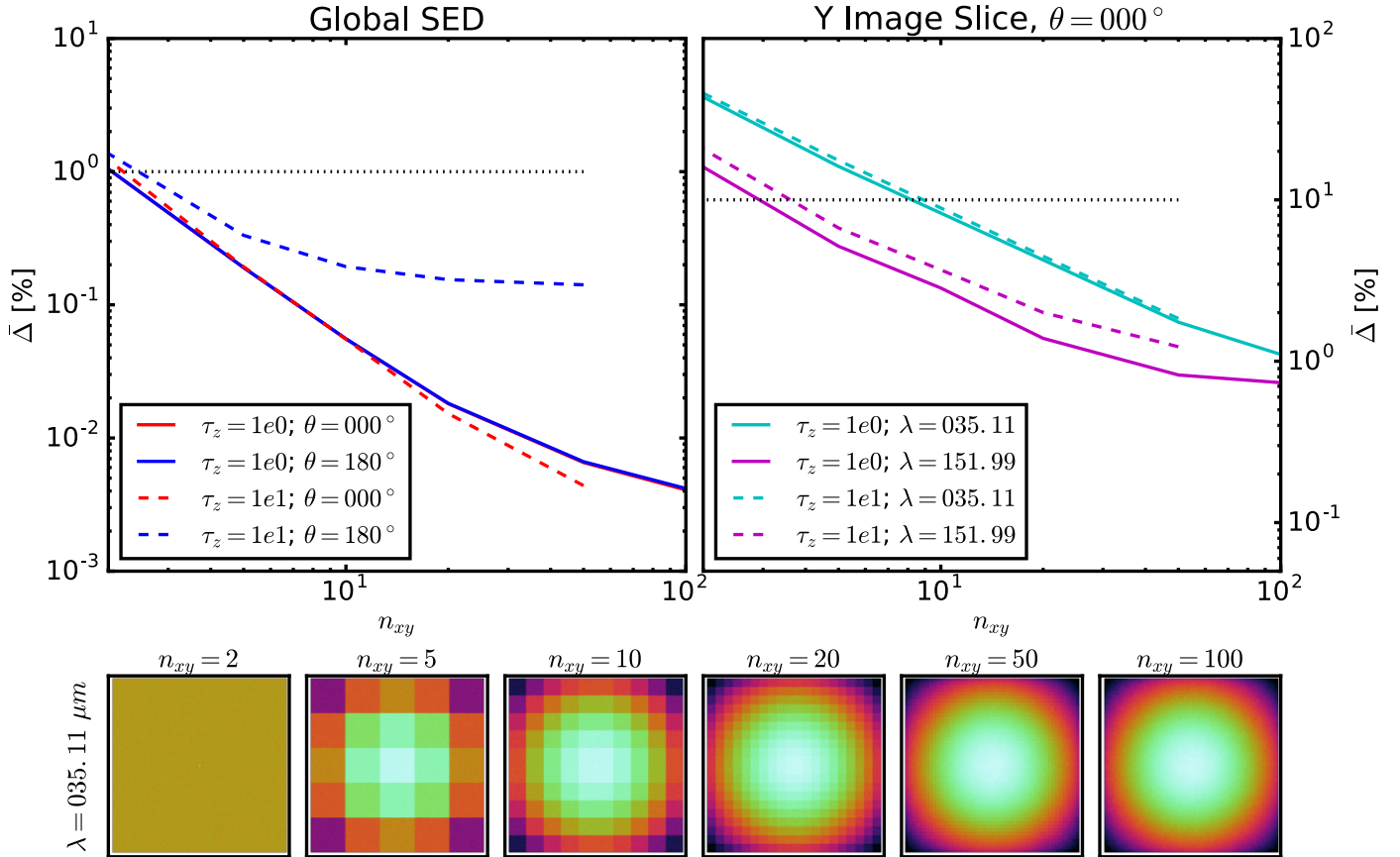
Fig. 14. Average deviations ($\bar{\Delta}$) versus n_{xy} are shown for the total global SEDs and Y image slices for $\tau_z(1\,\mu m) = 1$ and 10, and $\theta = 0°$ and 180°. The images are for the $\tau_z(1\,\mu m) = 1$ and $\theta = 0°$ case and illustrate the qualitative impact of increasing n_{xy}. The images are log scaled over the same range. The image Y slice results are shown only for $\theta = 0°$ as the results for $\theta = 180°$ are very similar. Only the image slices at two diagnostic infrared wavelengths probing the dust emission are shown as the ultraviolet and optical scattered light images are not sensitive to n_{xy} for Monte Carlo codes. The dashed and dotted horizontal lines give the 1% (global SED) and 10% (Y slice) lines. The results for $n_{xy} = 200$ ($\tau_z(1\,\mu m) \leq 1$) and $n_{xy} = 100$ ($\tau_z(1\,\mu m) = 10$) are not shown as $\bar{\Delta}$ is computed relative to these cases where $\bar{\Delta} = 0$ by definition.

on composite biasing was introduced to the Monte Carlo 3D dust RT community by Baes et al. (2016). The composite bias technique provides a way to sample two probability distributions while controlling the amplification of the resulting photon weight. Basically, the site of the next scattering is chosen from one of two different distributions with the frequency with which each distribution is used is controlled by the parameter ξ_{scat} that varies between 0 and 1. The first distribution is the standard $e^{-\tau}$ and the second is a much flatter distribution (e.g., a uniform distribution τ). For example, if $\xi_{scat} = 0.5$, then one half the time the first distribution is used and the other half the second is used. The weight of the photon is modified to account for the difference of the composite of the two distributions from the standard $e^{-\tau}$ distribution. The $\xi_{scat} = 0$ case corresponds to the standard $e^{-\tau}$ distribution.

The need for such a composite technique can be illustrated by considering the $\tau_z(1\,\mu m) = 10$ and $\theta = 90°$ case. For the standard scattering distribution ($\xi_{scat} = 0.0$), the back of the slab was approximately 10^{-100} fainter than the front of the slab for the 0.15 μm image. The approximate difference between the scattered light component from the front to back of the slab can be calculated by the simple approximation that intensity from singly scattered photons should be the albedo multiplied by $e^{-\tau}$. Assuming the scattered light at the front of the slab has $\tau = 0$ and the back of the slab has $\tau \sim 79$ (at 0.15 μm for $\tau_z(1\,\mu m) = 10$),

then the ratio should be $e^{-79} = 1.8 \times 10^{-35}$. This is much, much higher than seen using the standard scattering prescription.

Figure 18 illustrates the results for a range of ξ_{scat} at two representative wavelengths for the $\tau_z(1\,\mu m) = 10$ model. For $\xi_{scat} = 0.0$ it shows that most of the scattered photons are missing at the back of the slab. These scattered photons are computed with even small values ξ_{scat} as the interaction site for scattering is sampled from the composite function providing reasonable sampling at low *and* high optical depths. At $\xi_{scat} = 1.0$, a significant amount of non-Gaussian noise is seen at $\lambda = 151.99\,\mu m$ due to the extreme amplification of a small number of photons in the calculation. Setting $\xi_{scat} = 0.5$ provides a good compromise between sampling the low probability scattering events and controlling the amplification of the photon weight to be always less than two (Baes et al. 2016).

6. Discussion

For the $\tau_z(1\,\mu m) \leq 1$ cases, the results from the different codes agree within 0.3–2.9% for the global SEDs and within 3–11% for the Y slices. These are near or below the goal precisions for this benchmark.

For the $\tau_z(1\,\mu m) = 10$ case, the results from the different codes agree within 1.7–4.0% for the global SEDs, except for the scattered stellar component where the disagreement is up
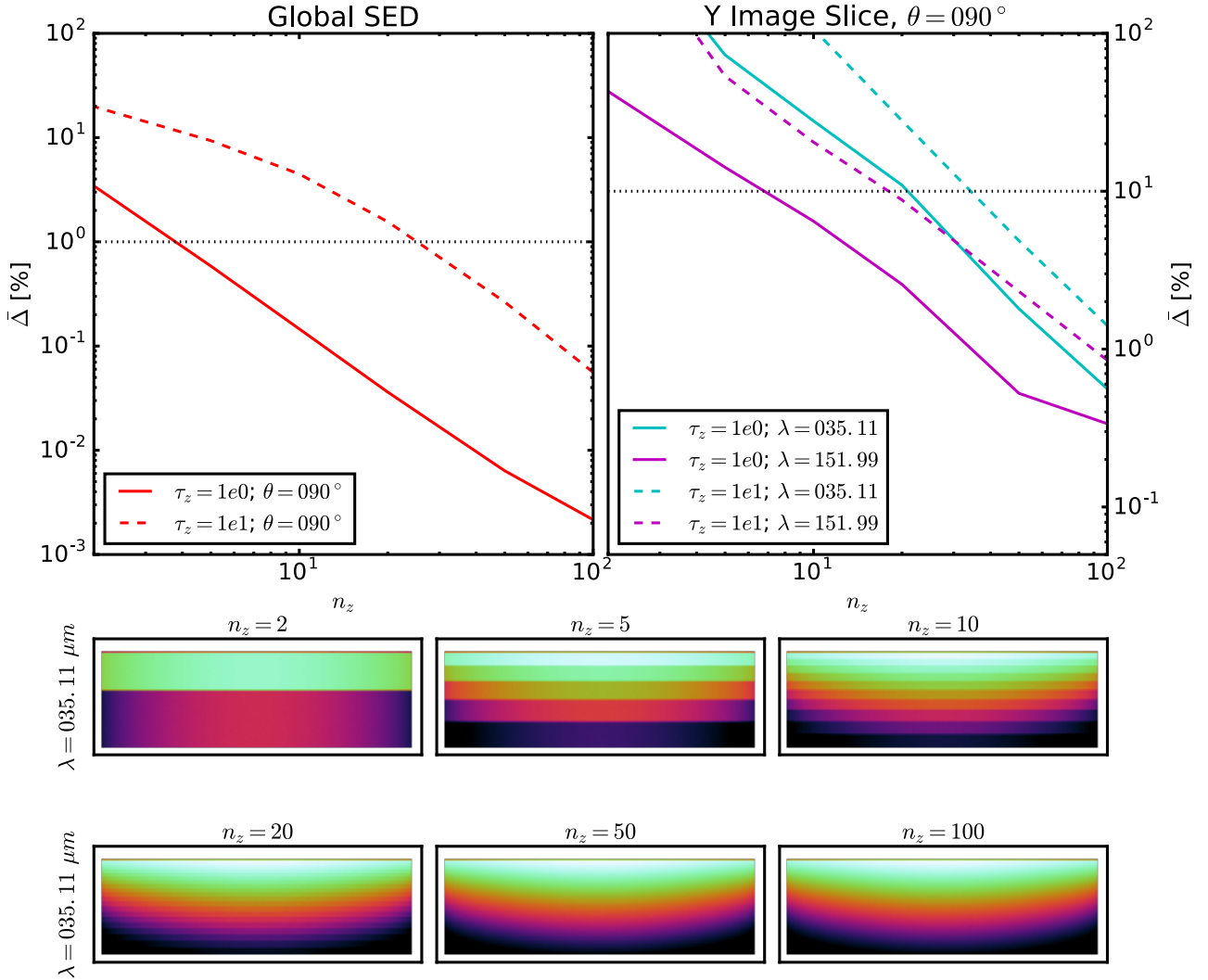
Fig. 15. Average deviations ($\bar{\Delta}$) versus n_z are shown for the total global SEDs and Y image slices for $\tau_z(1\ \mu m) = 1$ and 10 and $\theta = 90°$. The images are for the $\tau_z(1\ \mu m) = 1$ and $\theta = 90°$ case and illustrate the qualitative impact of increasing n_z. The images are log scaled over the same range. The DIRTY spatial grid uses a log spacing along the z axis and this is reflected in the images shown. Only the image slices at two diagnostic infrared wavelengths probing the dust emission are shown as the ultraviolet and optical scattered light images are not sensitive to n_z for Monte Carlo codes. The dashed and dotted horizontal lines give the 1% (global SED) and 10% (Y slice) lines. The results for $n_z = 200$ are not shown as $\bar{\Delta}$ is computed relative to this case where $\bar{\Delta} = 0$ by definition.

to 58%. The infrared Y slices agree within the goal precisions except for the sto case at 23.10 μm where the disagreement is 20%. The optical and UV Y slice deviations are very large, well beyond the goal precisions. These disagreements in the scattered flux are due to the continued challenge of performing accurate calculations at high optical depths. The most diagnostic viewing angle for these calculations is $\theta = 180°$ and it is striking that none of the seven codes produces equivalent results with systematic differences up to 10^4. While the disagreements are large for the scattered component at high optical depths, the scattered flux at high optical depths is not important as a heating term for the dust emission as evidenced by the agreement between codes for the IR emission SED and images (Fig. 10). The large differences for the mid-IR wavelengths for DART-Ray and TRADING are due to those codes not calculating the scattered component of the dust emission radiative transfer. In addition, some of the differences in the IR emission for DART-Ray are due to the lack of inclusion of dust self-heating.

The lack of agreement or convergence for the scattered component at high optical depths for dust RT codes is shown

quantitatively for the first time in this work. While it has been known for some time that high optical depths are challenging for dust RT codes, we have shown here that the main issue is with the scattered component and not the directly absorbed component. Previous benchmarks (Pascucci et al. 2004; Pinte et al. 2009) provided tests for high optical depths, but due to the 2D disk geometry used, they do not explicitly test the scattered component at such optical depths. Further, the 2D disk geometry in previous benchmarks had very high optical depths along the mid-plane of the disk, but very low optical depths along the rotation axis of the system. Thus, the global scattered flux from these benchmarks is dominated by scattering at low optical depths, not the high disk plane optical depths. The slab benchmark $\tau_z(1\ \mu m) = 10$ case with $\theta = 180°$ explicitly tests the scattered flux at very high optical depths as there are no paths for the scattered photons to the observer that do not go through at least the τ_z optical depth. Many of the codes in this benchmark have been run for the previous benchmarks and have given results that are consistent with the literature (e.g., Bianchi 2008; Robitaille 2011; Peest et al. 2017). For the reasons already stated, there
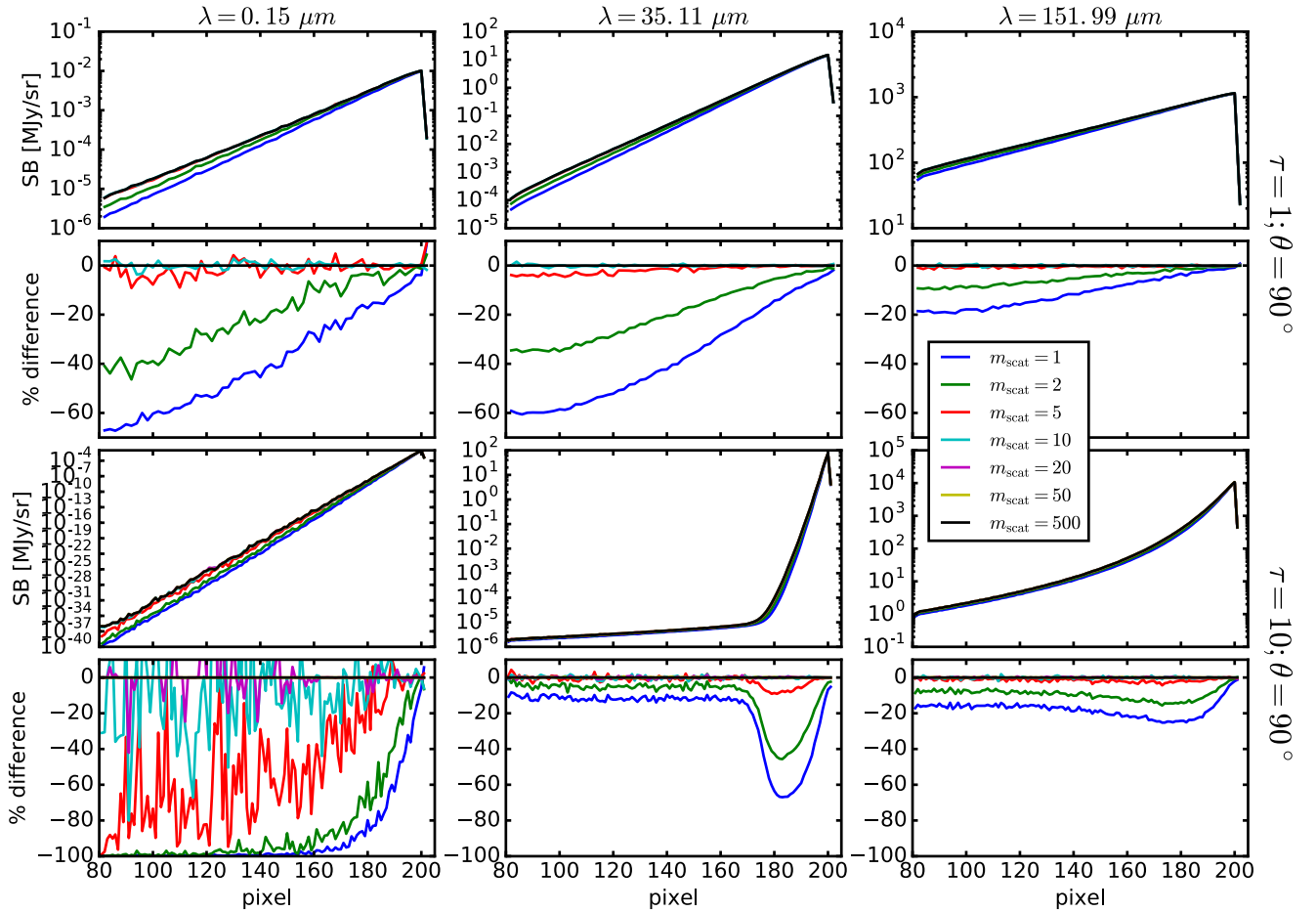
Fig. 16. The $\lambda = 0.15$ (*left*), 35.11 (*center*) and 151.99 μm (*right*) Y slices for $\theta = 90°$ and $\tau_z(1\,\mu$m$) = 1$ and 10 are plotted for a range of maximum allowed number of scatterings ($m_{\rm scat}$).

is no conflict between reproducing the previous benchmarks at high optical depths and the same codes not agreeing for the scattered component at high optical depths for the benchmark presented in this paper.

For the dust emission approximations, the codes agree best for the effective grain approximation (eff), with somewhat lower precision for the equilibrium-only approximation (equ), and worst for the full solution including stochastic emission (sto). The larger disagreement for the equ results may be due to the number of dust grain size bins adopted by each model and how the dust properties were averaged or interpolated for these bins. The larger disagreement for the sto results was expected given the different solution techniques used in the codes for the stochastically heated grains (Camps et al. 2015).

The inclusion of six Monte Carlo based radiative transfer codes provided a wealth of comparisons between codes based on the same technique. In addition, the inclusion of the DART-Ray code that is based on the alternative technique of Ray-Tracing allowed for comparisons to be made between the two techniques. This provided ample opportunity to find and remove bugs in all the codes. The solutions based on Monte Carlo and Ray-Tracing techniques were consistent, with the notable exception of the scattered component in the $\tau_z(1\,\mu$m$) = 10$ case.

6.1. Convergence tests

The convergence tests provide insight into the potential origin of the differences between codes when combined with Table 3.

The number of photons each model was run with (N – Table 3) proved sufficient to reach the desired precision in all cases except for stellar scattering for the $\tau_z(1\,\mu$m$) = 10$. Convergence testing indicates that, to reach the desired precision, the number of photon packets would exceed 10^9 with $\tau_z(1\,\mu$m$) = 10$. This is a potential cause of some of the large differences between models at UV/optical wavelengths at $\tau_z(1\,\mu$m$) = 10$, since some models were run with 10^8 photon packets. But this is not the only cause, as targeted additional tests of just the UV scattering for the $\tau_z(1\,\mu$m$) = 10$ case with many more than 10^9 photons do not show convergence. The number of x and y bins (n_{xy}) are sufficient in all the model runs. The number of z bins (n_z) are sufficient for all but the DART-Ray results for the $\tau_z(1\,\mu$m$) = 10$ case. The maximum number of scatterings ($m_{\rm scat}$) is of the order of 20 with most codes computing many more scatterings, with the exception of DART-ray and (marginally) SOC. The maximum number of dust heating iterations ($m_{\rm iter}$) is sufficient for all the codes except for DART-Ray for the $\tau_z(1\,\mu$m$) = 10$ case.

6.2. Lessons learned

To very carefully setup, specify, and define parameters as much as possible in order to ensure that all the codes perform the same calculation was found to be critical. One area that was found to be important was to clearly specify the wavelength grid and ensure that each code performed calculations at the defined wavelengths. Additionally, the normalization of the slab optical
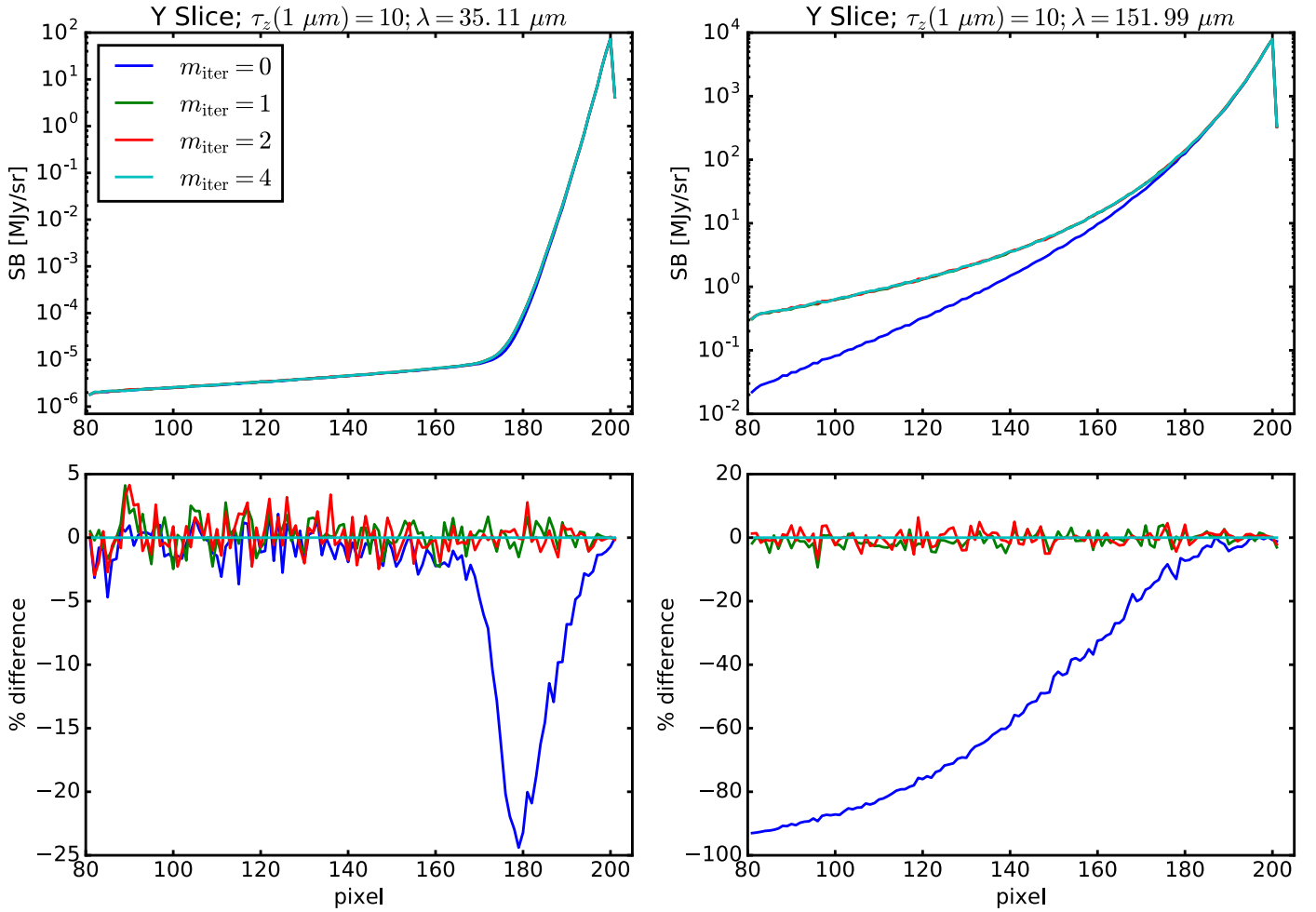
Fig. 17. The λ = 35.11 (*left*) and 151.99 μm (*right*) Y slices for $\tau_z(1\,\mu$m) = 10 and θ = 90° are plotted for a range of dust self-heating iterations ($m_{\rm iter}$).

depth was initially at 0.55 μm, but was changed to be exactly at one of the specified grid points (1 μm) to avoid, as much as possible, interpolation errors in the normalization. It was also critical to get all the results of the models into the same format and orientations, something that took a surprising amount of time due to the different assumptions made in the different codes. We also spent significant time establishing a common terminology for different parts of the models and benchmark.

Another lesson learned was that there were minor bugs and/or different conventions that the initial comparisons revealed. Most of the participating codes made improvements as a result. These improvements included removing minor bugs that did not significantly change the results but did improve the ability to compare different codes. Some of these bugs were revealed due to the large parameter space explored by this benchmark, often beyond the range that had been tested in the codes previously.

A major lesson learned was that the codes had significant difficulty with the scattered component for the $\tau_z(1\,\mu$m) = 10 case. This case pushed the codes into an area where dust scattering is critical, both for the stellar and dust emission photons. The initial results revealed approximately 10^{60} differences at the back of the slab in the stellar scattered light for the λ = 0.15 μm Y slice (Sect. 5.5). In particular, the Ray-Tracing (DART-ray) results were much higher than many of the Monte Carlo results. As Monte Carlo codes have a known limitation in not fully probing the high optical depths, this was not particularly

surprising. Test cases with some of the Monte Carlo codes at this wavelength with more photons did show smaller differences, but these were still very large. Additional test cases were run allowing for larger numbers of scatterings and these also showed smaller differences, but again these were still relatively large. These differences motivated the inclusion of the composite biasing technique as part of all of the Monte Carlo codes to better probe the high optical depth scatterings (Baes et al. 2016). While the codes still have significant differences for this case and wavelength, they are much smaller at only $\sim$10^4. Clearly more work is needed to understand the origin of these differences. This work has started and has indicated that the origin may be related to very low probability multiple scattering events that dominate the scattered light images at these high optical depths. While under active investigation, the solution to this issue is beyond the scope of this paper, however this problem did reveal that there may be other very low probability parts of parameter space that are not probed well. An example of this is the dust emission when it varies strongly over the model grid. This is the case for mid-IR wavelengths for $\tau_z(1\,\mu$m) = 10 where the dust emission at the front of the slab is many orders of magnitude larger than that from the back of the slab. Using the dust emission as the probability for emitting a mid-IR photon can lead to very few photons being emitted in the back of the slab. Of course increasing the number of photons in a run will alleviate this issue, but at the expense of longer run times. The
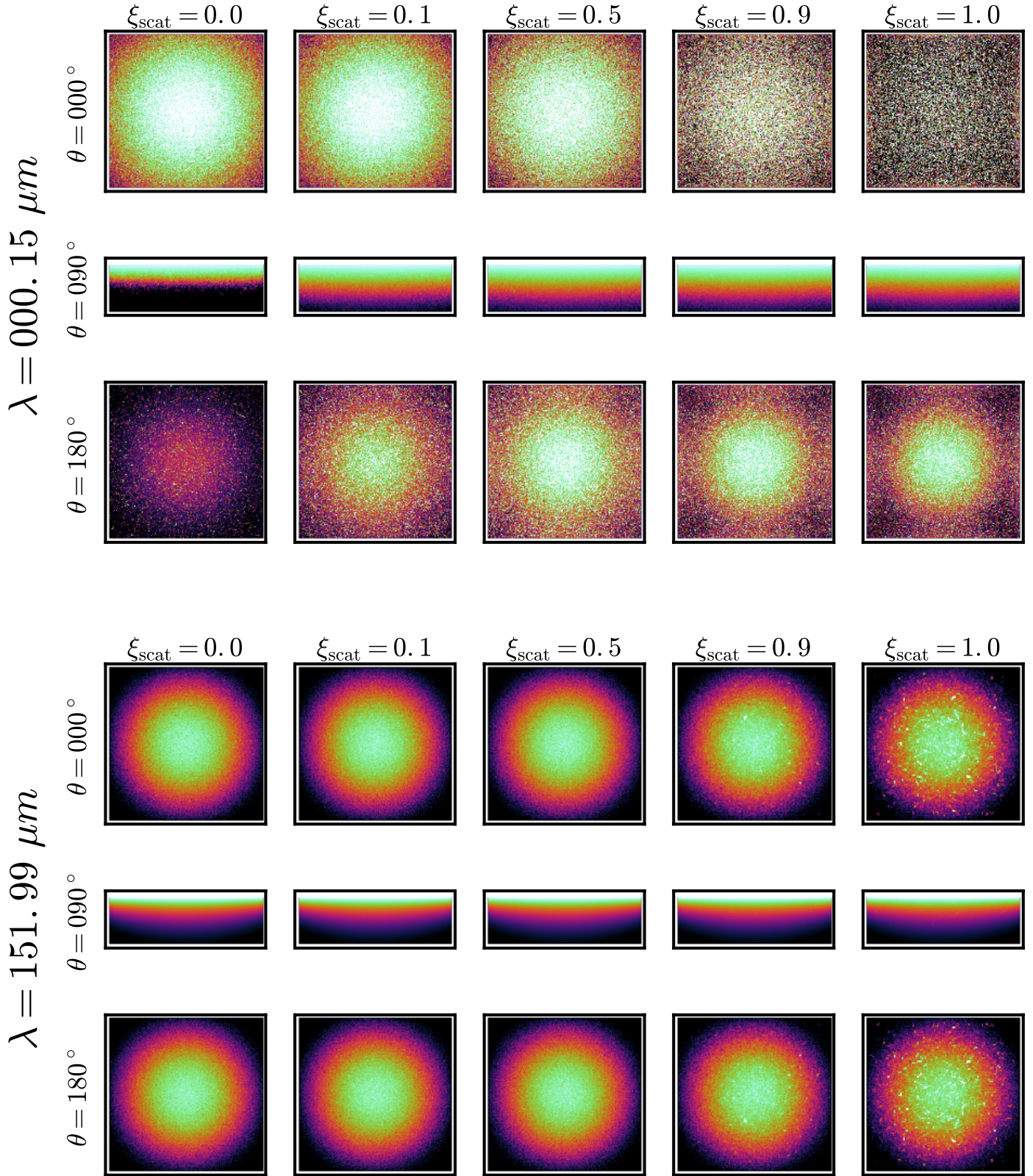
Fig. 18. Images in the UV and IR are shown for $\tau_z(1 \mu m) = 10$, $\theta = 0°$, $90°$, and $180°$, and a range of ξ_{scat} values. The images are log scaled and share the same scaling at each unique combination of λ and θ.

composite biasing technique can be used to provide for better sampling of the spatial dust emission by, for example, emitting photons one half of the time sampled from the dust emission and one half of the time sampled uniformly in the slab. This was implemented in the DIRTY code producing much better sampled IR images for the same number of photons emitted. Another solution may be to use the Bjorkman & Wood (2001) instantaneous emission technique where the spatial locations of the dust

emission photons are based on the locations of dust absorptions, but possibly at the expense of longer run times (Baes et al. 2005; Chakrabarti & Whitney 2009).

One minor lesson learned was that it would be useful to have the convergence tests run prior to having all the codes run the benchmark. The convergence tests for this benchmark were generally run after the comparison of results from all the codes was well underway. While none of the convergence tests revealed that the codes needed to rerun the benchmarks, we did discover that the BASIC wavelength grid for the eff case was close to being too coarse for our goal precision. The impact of the wavelength grid resolution appeared in the dust emission spectrum with higher resolution grids showing slightly hotter dust. Convergence tests on the wavelength grid resolution were done with multiple codes for the eff case showing that the global SED convergence was near 1% for the adopted BASIC grid for the $\tau_z(1\,\mu m) = 0.01$ case. The global SED convergence precision improved for higher $\tau_z(1\,\mu m)$ values. While this is within the goal precision, in hindsight we likely would have adopted a finer resolution grid to ensure that this issue was well below the goal precision of 1%.

We all learned that there is no "easy" benchmark. The expectation by many of us was that the slab benchmark would be straightforward and take relatively little time to complete. This was not the case and many of us found this geometrically simple benchmark to be deceptively complex. This benchmark was useful not only for debugging and refining our codes, but also for deepening our understanding of dust radiative transfer in general.

Many of the lessons learned during the work on this benchmark should be applicable to other 3D dust radiative transfer geometries. The lack of agreement between the codes for the scattered light at high optical depths ($\tau > 10$) calls for continued caution in interpreting results obtained from any code. The importance of the convergence tests in illuminating the importance of different parameters in the precision of the solution for this benchmark illustrates that such convergence tests should be done for all geometries. Convergence tests can be done by anyone using a dust radiative transfer code (not just the coders), and will provide confidence in the results as well as a deeper understanding of the complex dust radiative transfer problem.

7. Summary

We present the first 3D dust radiative transfer benchmark. This benchmark is composed of a rectangular slab of constant density dust externally illuminated by a hot, UV-bright star. The cases in this benchmark include optical depths from $\tau_z(1\,\mu m) = 0.01$–10 and three different dust emission assumptions (effective grain, equilibrium-only grains, and the full solution including stochastically heated grains). Results from seven codes are presented, six based on Monte Carlo techniques and one based on Ray-Tracing. The results are given as global SEDs and images at selected wavelengths for a range of viewing angles. The results are in good agreement for $\tau_z(1\,\mu m) \leq 1$ with precisions of ~1% for the global SEDs and $\leq 10\%$ for slices through the images. The results are in good agreement for the $\tau_z(1\,\mu m) = 10$ case, except for the stellar scattered component. The setup of this benchmark to explicitly probe the components of the dust radiative transfer problem allowed us to quantify the lack of agreement for the scattered component at high optical depths. This work provides a benchmark for future 3D RT codes and illustrates remaining challenges for 3D dust RT in the very optically thick regime.

References

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Baes, M., & Camps, P. 2015, Astronomy and Computing, 12, 33
Baes, M., & Dejonghe, H. 2002, MNRAS, 335, 441
Baes, M., Stamatellos, D., Davies, J. I., et al. 2005, New Astron., 10, 523
Baes, M., Fritz, J., Gadotti, D. A., et al. 2010, A&A, 518, L39
Baes, M., Verstappen, J., De Looze, I., et al. 2011, ApJS, 196, 22
Baes, M., Gordon, K. D., Lunttila, T., et al. 2016, A&A, 590, A55
Balog, Z., Müller, T., Nielbock, M., et al. 2014, Exp. Astron., 37, 129
Bianchi, S. 2007, A&A, 471, 765
Bianchi, S. 2008, A&A, 490, 461
Bianchi, S., & Xilouris, E. M. 2011, A&A, 531, L11
Bianchi, S., Ferrara, A., & Giovanardi, C. 1996, ApJ, 465, 127
Bianchi, S., Davies, J. I., & Alton, P. B. 2000, A&A, 359, 65
Bjorkman, J. E., & Wood, K. 2001, ApJ, 554, 615
Bohlin, R. C., Gordon, K. D., Rieke, G. H., et al. 2011, AJ, 141, 173
Calzetti, D., Bohlin, R. C., Gordon, K. D., Witt, A. N., & Bianchi, L. 1995, ApJ, 446, L97
Camps, P., & Baes, M. 2015, Astronomy and Computing, 9, 20
Camps, P., Baes, M., & Saftly, W. 2013, A&A, 560, A35
Camps, P., Misselt, K., Bianchi, S., et al. 2015, A&A, 580, A87
Chakrabarti, S., & Whitney, B. A. 2009, ApJ, 690, 1432
De Geyter, G., Baes, M., Camps, P., et al. 2014, MNRAS, 441, 869
De Looze, I., Baes, M., Fritz, J., & Verstappen, J. 2012, MNRAS, 419, 895
Deschamps, R., Braun, K., Jorissen, A., et al. 2015, A&A, 577, A55
Gordon, K. D., Witt, A. N., Carruthers, G. R., Christensen, S. A., & Dohne, B. C. 1994, ApJ, 432, 641
Gordon, K. D., Calzetti, D., & Witt, A. N. 1997, ApJ, 487, 625
Gordon, K. D., Clayton, G. C., Witt, A. N., & Misselt, K. A. 2000, ApJ, 533, 236
Gordon, K. D., Misselt, K. A., Witt, A. N., & Clayton, G. C. 2001, ApJ, 551, 269
Hendrix, T., Keppens, R., & Camps, P. 2015, A&A, 575, A110
Holwerda, B. W., Bianchi, S., Böker, T., et al. 2012, A&A, 541, L5
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90
Ivezic, Z., Groenewegen, M. A. T., Men'shchikov, A., & Szczerba, R. 1997, MNRAS, 291, 121
Johnston, K. G., Robitaille, T. P., Beuther, H., et al. 2015, ApJ, 813, L19
Jonsson, P. 2006, MNRAS, 372, 2
Juvela, M. 2005, A&A, 440, 531
Juvela, M., & Padoan, P. 2003, A&A, 397, 201
Juvela, M., Pelkonen, V.-M., Padoan, P., & Mattila, K. 2006, A&A, 457, 877
Karl, S. J., Lunttila, T., Naab, T., et al. 2013, MNRAS, 434, 696
Koepferl, C. M., Robitaille, T. P., Morales, E. F. E., & Johnston, K. G. 2015, ApJ, 799, 53
Kuiper, R., Klahr, H., Dullemond, C., Kley, W., & Henning, T. 2010, A&A, 511, A81
Law, K.-H., Gordon, K. D., & Misselt, K. A. 2011, ApJ, 738, 124
Lewis, N. K., Cook, T. A., Wilton, K. P., et al. 2009, ApJ, 706, 306
Lunttila, T., & Juvela, M. 2012, A&A, 544, A52
Malinen, J., Juvela, M., Collins, D. C., Lunttila, T., & Padoan, P. 2011, A&A, 530, A101
Misselt, K. A., Gordon, K. D., Clayton, G. C., & Wolff, M. J. 2001, ApJ, 551, 277

Narayanan, D., Turk, M., Feldmann, R., et al. 2015, Nature, 525, 496

Natale, G., Popescu, C. C., Tuffs, R. J., & Semionov, D. 2014, MNRAS, 438, 3137

Natale, G., Popescu, C. C., Tuffs, R. J., et al. 2015, MNRAS, 449, 243

Pascucci, I., Wolf, S., Steinacker, J., et al. 2004, A&A, 417, 793

Peest, C., Camps, P., Stalevski, M., Baes, M., & Siebenmorgen, R. 2017, A&A, 601, A92

Pierini, D., Gordon, K. D., Witt, A. N., & Madsen, G. J. 2004, ApJ, 617, 1022

Pinte, C., Harries, T. J., Min, M., et al. 2009, A&A, 498, 967

Robitaille, T. P. 2011, A&A, 536, A79

Robitaille, T. P., & Whitney, B. A. 2014, in The Labyrinth of Star Formation, eds. D. Stamatellos, S. Goodwin, & D. Ward-Thompson, Astrophys. Space Sci. Proc., 36, 157

Robitaille, T. P., Churchwell, E., Benjamin, R. A., et al. 2012, A&A, 545, A39

Saftly, W., Camps, P., Baes, M., et al. 2013, A&A, 554, A10

Saftly, W., Baes, M., & Camps, P. 2014, A&A, 561, A77

Saftly, W., Baes, M., De Geyter, G., et al. 2015, A&A, 576, A31

Schneider, R., Bianchi, S., Valiante, R., Risaliti, G., & Salvadori, S. 2015, A&A, 579, A60

Sipilä, O., Harju, J., & Juvela, M. 2011, A&A, 535, A49

Stalevski, M., Fritz, J., Baes, M., Nakos, T., & Popović, L. Č. 2012, MNRAS, 420, 2756

Steinacker, J., Baes, M., & Gordon, K. D. 2013, ARA&A, 51, 63

Vijh, U. P., Witt, A. N., & Gordon, K. D. 2003, ApJ, 587, 533

Witt, A. N., & Gordon, K. D. 1996, ApJ, 463, 681

Witt, A. N., & Gordon, K. D. 2000, ApJ, 528, 799

Ysard, N., Juvela, M., & Verstraete, L. 2011, A&A, 535, A89

Zubko, V., Dwek, E., & Arendt, R. G. 2004, ApJS, 152, 211