Mon. Not. R. Astron. Soc. 272, 291–307 (1995)

A large-particle Monte Carlo code for simulating non-linear high-energy processes near compact objects

Boris E. Stern,[1]★ Mitchell C. Begelman,[2,3]† Marek Sikora[2,4]‡ and Roland Svensson[5]§

[1] *Institute for Nuclear Research, Russian Academy of Sciences, Moscow 117312, Russia*
[2] *Joint Institute for Laboratory Astrophysics, University of Colorado and National Institute of Standards and Technology, Campus Box 440, Boulder, CO 80309-0440, USA*
[3] *Department of Astrophysical, Planetary and Atmospheric Sciences, University of Colorado, Boulder, CO 80309, USA*
[4] *N. Copernicus Astronomical Center, Bartycka 18, 00-716 Warsaw, Poland*
[5] *Stockholm Observatory, S-133 36, Saltsjöbaden, Sweden*

Accepted 1994 August 1. Received 1994 July 26; in original form 1994 April 19

ABSTRACT

High-energy radiation processes in compact cosmic objects are often expected to have a strongly non-linear behaviour. Such behaviour is shown, for example, by electron–positron pair cascades and the time evolution of relativistic proton distributions in dense radiation fields. Three independent techniques have been developed to simulate these non-linear problems: the kinetic equation approach; the phase-space density (PSD) Monte Carlo method; and the large-particle (LP) Monte Carlo method.

In this paper, we present the latest version of the LP method and compare it with the other methods. The efficiency of the method in treating geometrically complex problems is illustrated by showing results of simulations of 1D, 2D and 3D systems. The method is shown to be powerful enough to treat non-spherical geometries, including such effects as bulk motion of the background plasma, reflection of radiation from cold matter, and anisotropic distributions of radiating particles. It can therefore be applied to simulate high-energy processes in such astrophysical systems as accretion discs with coronae, relativistic jets, pulsar magnetospheres and gamma-ray bursts.

Key words: plasmas – radiation mechanisms: non-thermal – radiative transfer – methods: numerical – galaxies: jets – gamma-rays: bursts.

1 INTRODUCTION

The spectral and temporal properties of X-ray and gamma-ray emission from many cosmic compact objects suggest the involvement of non-thermal radiation processes. Non-thermal processes give rise to the radiation from relativistic jets in active galactic nuclei (AGNs; see, e.g., review by Sikora 1994), from pulsars (Daugherty & Harding 1982; Chen & Ruderman 1993), and, possibly, from gamma-ray bursts (Rees & Mészáros 1992; Mészáros & Rees 1993). They may also occur in accretion disc coronae in galactic X-ray binaries (Lingenfelter & Hua 1991; Sunyaev et al. 1992) and in AGNs (Zdziarski, Lightman & Maciołek-Niedźwiecki 1993; Madejski et al. 1994). The interpretation of observed radiation spectra from these sources is not unique, however, and the acceleration mechanisms of relativistic particles are unknown. One way to obtain a better understanding of the physical nature of these objects is to perform numerical simulations of the radiation processes in different scenarios and to compare the results with observations.

Unfortunately, this task is not easy since the non-thermal processes in highly compact objects can be very non-linear. For example, consider a type of relativistic particle that radiates due to interactions with photons (e.g. Comptonization by electrons or photomeson production by protons). The particles can interact with 'external' (ambient) photons, as well as with 'internally' generated photons produced either by the particles themselves or by cascade processes initiated by the particles. Furthermore, the number density of radiating particles can be strongly increased by pair production resulting from absorption of gamma-rays by X-rays, with these photons themselves being produced by relativistic pairs.

Non-linear radiation processes have been simulated successfully by solving a system of kinetic equations (Fabian et al. 1986; Kirk & Mastichiadis 1989; Coppi 1992; Coppi, Blandford & Rees 1993). The kinetic equation approach

★ Internet: stern@inr.msk.su
† Internet: mitch@jila.colorado.edu
‡ Internet: sikora@camk.edu.pl
§ Internet: svensson@astro.su.se

readily achieves good particle statistics and is very efficient for geometrically simple models. However, the efficiency of the method falls dramatically with the increasing dimensionality of the system (Xu 1994), because it requires discretization of phase space as well as computations of interactions between each pair of momentum cells. In addition, the end products of the interactions must be distributed among many momentum cells within the allowed kinematic limits.

Monte Carlo methods can be much more economical as regards the number of computational steps, since it is not necessary to calculate all possible interactions among particles in different regions of phase space. Instead, one selects and follows specific particles in order to characterize the overall distributions. The trick is then to find the best way to represent the particle distributions statistically. To date, two approaches have been used to accomplish this: the phase-space density (PSD) array representation and the large-particle (LP) representation.

In the PSD approach, the system is represented in the same way as in the kinetic equation method – by densities of interacting particles in cells of a discretized phase space. The transition of a particle between spatial cells is simulated using an escape probability formalism, and particle transitions in momentum space due to interactions are simulated using standard Monte Carlo methods. So far, this approach has been used only to simulate homogeneous spherically symmetric systems (Stern 1985; Xu 1994).

In the LP approach, the system is represented by an array of 'large particles', each of which corresponds to a group of real particles sharing the same physical parameters – particle type, position and momentum vector. Compared to the PSD approach, the LP method uses a more flexible representation of the particle density in phase space. Associated with each LP is a statistical weight, which is proportional to the number of real particles represented by the LP. Adjustment of the weight allows one to improve the statistical representation in those regions of phase space that contain relatively fewer particles. Such a representation can be especially useful in non-thermally radiating systems, where sparsely populated regions of phase space may contain most of the energy. This type of representation was used to model non-linear systems with Coulomb-type interactions, such as gravitating systems or charged particle beams in accelerators. To treat non-linear radiation processes near cosmic compact objects using this approach, it has been necessary to invent new techniques. Such techniques were developed by Stern (1988) for pure electromagnetic cascades, and by Stern, Svensson & Sikora (1990) and Stern, Sikora & Svensson (1992) for hadronic–electromagnetic cascades.

In this paper we describe the application of the LP method to non-linear pair cascades. The LP method has several important advantages over other methods: (i) particle transport is treated naturally, without the use of escape probabilities; (ii) the efficiency of the code depends very weakly on the dimensionality and geometric complexity of the problem; and (iii) the representation given by large-particle arrays does not depend on model features such as the geometry, the magnetic field structure or the dynamical situation. A code based on this representation is easily programmed. The LP method is described and compared with other methods in Section 2. In Section 3, we perform simulations of non-linear radiation processes, both to test the code by comparing with earlier work and to show the power of the LP method in solving problems never treated before. Finally, in Section 4, we discuss and summarize our results.

2 THE LARGE-PARTICLE METHOD

Before introducing the LP method it is worth explaining the difference between linear and non-linear Monte Carlo simulations. In a linear Monte Carlo method the objects of a simulation are test particles interacting only with a background medium, not with each other. The parameters of the medium are fixed or updated gradually, taking into account the cumulative effects of the incident particles. Since the test particle histories are considered to be independent of one another, there is no reason to store information about all particle trajectories at the same moment. A well-known application of a linear Monte Carlo approach in astrophysics is the work of Pozdnyakov, Sobol & Sunyaev (1977), which deals with photon Comptonization. In non-linear situations such as those described in the Introduction, however, it is necessary to treat the interactions between simulated particles. This means that information about all interacting particles must be stored simultaneously. We will argue in this paper that the LP method is a very efficient technique for organizing and storing this information.

In the LP method, the non-linear system of interacting particles is represented by an array of large particles (LPs). Each LP is tagged by the type, energy, location in space and direction of motion of the physical particles it represents, plus a statistical weight ω, which is proportional to the number of physical particles represented by the LP. Like the physical particles, LPs successively interact with one another, changing their attributes and moving in space between interactions. Because the evolution of all particles must be synchronized in time, time-steps for the simulation must be sufficiently short that changes in the particle populations during one time-step are small. We now discuss some important techniques used in the LP method: operations with statistical weights; creation and destruction of LPs; calculation of the mean free time of an LP between interactions; choice of a target LP; and modifications of the method in cases where the straightforward LP technique is inefficient.

2.1 Statistical weights

The number of particles in the low-energy range is typically several orders of magnitude larger than in the high-energy range. Therefore LPs representing particles of different energies must have different statistical weights. The LP energy weight function for particles of type i is defined to be

$$E_i(\varepsilon) \equiv \omega_i(\varepsilon)\,\varepsilon, \tag{1}$$

where ω_i is the statistical weight and ε is the energy of the physical particle represented by this LP. Below we refer to ε as the LP energy, whereas E_i represents the total energy of all particles represented by the LP.

In general, E_i can be an arbitrary function of energy, different for different kinds of particles. The simplest approach is to set $E_i(\varepsilon) = \text{constant}$, which means that all LPs share the energy of the system equally. The advantages of choosing such an energy weight function are that the parts of the particle distribution containing most of the energy are best

represented statistically, and the number of LPs is conserved for a system with fixed total energy.

An energy-dependent $E_i(\varepsilon)$ can, however, be useful; for example, if the simulated system includes a luminous and stable feature like a blackbody UV 'bump' which dominates over other parts of the spectrum (as in AGN spectra). In this case, we can reduce the number of 'blackbody' LPs by using a larger E_i in the spectral region around the blackbody maximum. To cite another example, suppose the spectral radiation flux in the system, F_ε, is approximately a power-law function with a spectral flux index α (i.e. $F_\varepsilon \propto \varepsilon^{-\alpha}$). Here, the choice of $E_i \propto \varepsilon^{1-\alpha}$ ensures uniform energy resolution over the spectrum.

2.2 Creation and destruction of LPs

Since the number of physical particles represented by an LP depends on particle energy ε (and possibly on time), the statistical weight of an LP can change with time, and one has to be careful that particle numbers are conserved appropriately. Particle number conservation is achieved in a time-averaged sense, in terms of the expectation values of the particle numbers represented by an LP before and after a collision. To accomplish this, an LP is either killed or reproduced in a number of identical copies with some probability, following each interaction.

Consider a large particle LP_1, with statistical weight ω_1, which changes energy from ε_1 to ε_2 (without changing particle type) due to an interaction. One can say that LP_1 is replaced by a new large particle, which we denote by LP_2. According to equation (1), the statistical weight of LP_2 is given by

$$\omega_2 = \omega_1 \frac{\varepsilon_1}{\varepsilon_2} \frac{E_i(\varepsilon_2)}{E_i(\varepsilon_1)}; \tag{2}$$

i.e. in general LP_2 will represent a *different number* of physical particles than LP_1. Particle number conservation requires that the change of the statistical weight be compensated by a change in the number of LPs representing the same group of physical particles. The mean number of copies of LP_2 produced in the interaction must be

$$n_2 \equiv \frac{\omega_1}{\omega_2} = \frac{\varepsilon_2}{\varepsilon_1} \frac{E_i(\varepsilon_1)}{E_i(\varepsilon_2)}, \tag{3}$$

which generally is not an integer. Therefore we enforce particle conservation in a statistical sense, by killing the LP_2 particle with probability $1-n_2$ when $n_2<1$, and by copying it (on average) n_2 times when $n_2>1$. For a constant energy weight function, the two cases correspond, respectively, to energy loss and energy gain by the physical particles represented by LP_1.

If a new particle of type j, represented by LP_3, is produced due to an interaction, its expectation statistical weight, ω_3, must satisfy

$$n_3 \equiv \frac{\omega_1}{\omega_3} = \frac{\varepsilon_3}{\varepsilon_1} \frac{E_i(\varepsilon_1)}{E_j(\varepsilon_3)}, \tag{4}$$

where the energy of the new particle, ε_3, is determined by the kinematics of the interaction. For $n_3<1$, the new particle is produced with a probability n_3, otherwise it is produced in n_3 copies on average.

In practice we will consider mainly two-body interactions, in which one particle (e.g. LP_1) can be considered the *incident* particle and the other (e.g. LP_4) the *target* particle. In order to avoid double counting, we take into account only interactions for which the incident particle has a lower statistical weight than the target particle, i.e. $\omega_1<\omega_4$. Since LP_4 represents a larger number of physical particles than LP_1, only a fraction ω_1/ω_4 of LP_4 is 'used up' in a given interaction. To conserve target particle number in a statistical sense, LP_4 has a probability $1-(\omega_1/\omega_4)$ of surviving an interaction.

Finally, the energy weights of all LPs have a common factor which is variable with time, if the total energy of the system is variable. This ensures that the number of LPs does not exceed the maximum permitted in the array, while maintaining enough LPs to obtain close to optimal statistics. For example, if the number of required LPs exceeds some technical limit, the statistical weight of all LPs can be multiplied by a common factor $f>1$, and each LP is then killed with probability $1-f^{-1}$.

2.3 Maximum cross-section method

In order to follow the evolution of a given incident LP, we must determine its mean free time for interacting with another LP and choose a partner from among possible target LPs. This could be done by implementing the following procedure.

(1) Calculate the probabilities for the incident LP to interact with all other LPs contained in the volume V around it, taking into account any energy and angular dependence but assuming that the spatial distribution of particles in the region is uniform. These probabilities are proportional to the partial interaction rates, $P_{it}=\omega_t\tilde{\sigma}_{it}(\varepsilon_i,\varepsilon_t,\theta)/V$, where the index '$i$' refers to the incident LP, the index 't' refers to the target LP, θ is the angle of interaction, and $\tilde{\sigma}_{it}\equiv\sigma_{it}v_{it}$ is the total cross-section multiplied by the relative velocity.

(2) Calculate the sum $P_i=\sum_t P_{it}$ and sample the mean free time using the standard Monte Carlo procedure,

$$t = -\ln(\xi)/P_i, \tag{5}$$

where ξ is a random number between 0 and 1.

(3) Choose the type of interaction, the type of target particle, and the actual target particle, using the partial interaction probabilities.

(4) Simulate the interaction.

This procedure, while clear and straightforward, is very inefficient. It requires $O(N)$ computational steps (where N is the total number of LPs) per time-step to simulate the evolution of each incident LP. Thus the total system simulation time scales as $O(N^2)$.

The efficiency can be improved dramatically by adopting a *virtual process* technique. The trick is to replace the true cross-section for each type of interaction l, $\sigma^{(l)}_{\rm real}$, with a maximum total cross-section (including all possible types of interaction), $\sigma^{(l)}_{\max}$, and to assume that $v_{it}=2c$. In this method, the maximum partial interaction rates, $P^{(l)}_{it,\max}=\omega_t\tilde{\sigma}^{(l)}_{\max}/V$, depend only on ω_t and do not depend on the properties of the incident LP. We define the LP statistical weight cumula-

tive arrays $S(j)=\sum_{k=1,j}\omega_k/V$, where j denotes the LPs in order of decreasing energy. The arrays are calculated once at the beginning of each time-step, for each type of particle. Now, the mean free time can be sampled from equation (5), but with P_i being replaced by $P_{i,\max}=\sum_l P_{i,\max}^{(l)}$, where

$$P_{i,\max}^{(l)}=\tilde{\sigma}_{\max}^{(l)}[S(j_{\max})-S(j_{\min}-1)], \tag{6}$$

where $j_{\min}$ and $j_{\max}$ correspond to the minimum and maximum energies of LPs that can participate in the interaction as a target [$j_{\min}=1$ if there is no energy threshold, and $S(0)\equiv 0$ has to be put in equation (6)]. After the type of target particle and type of interaction have been chosen, the index j of a possible target LP is searched for from the condition

$$S(j)-S(j_{\min}-1)>\xi[S(j_{\max})-S(j_{\min}-1)] > S(j-1)-S(j_{\min}-1). \tag{7}$$

After a tentative target, LP_j, has been found and all parameters of the interacting particles are known, it is possible to calculate the true cross-section, $\tilde{\sigma}_{\text{real}}$, and the relative velocity, v_{it}, of the interaction. This tentative interaction has a probability $\sigma_{\text{real}}v_{it}/\tilde{\sigma}_{\max}$ of being accepted. If it is accepted, the interaction is simulated with a standard Monte Carlo technique, and the parameters of the LPs are changed accordingly; otherwise, the interaction is cancelled and the incident LP proceeds on its trajectory without changing parameters.

The maximum cross-section method can easily be understood in the following way. Let us introduce a *virtual* channel for an interaction that (i) has a cross-section σ_{virt}; (ii) does not change the parameters of the interacting particles; and (iii) satisfies $(\sigma_{\text{real}}+\sigma_{\text{virt}})v_{it}=\sigma_{\max}2c$. The total interaction rate is then proportional to $\sigma_{\max}$, and, after sampling the mean free time with $\tilde{\sigma}_{\max}$, we find that the probability of choosing the *real* channel is $\sigma_{\text{real}}v_{it}/\tilde{\sigma}_{\max}$. This is the probability of a real interaction mentioned above.

Now, let us evaluate the simulation efficiency using this method. The number of LPs, N, that we deal with should be large enough to avoid large statistical fluctuations. The number of computational steps for pre-calculating the cumulative arrays scales as $O(N)$. Since the procedure of finding a target LP from equation (7) requires $O[\log(N)]$ steps for each time-step, the total time required for the simulation scales as $O[N\log(N)]$. The reduction in the number of computational steps from $O(N^2)$ to $O[N\log(N)]$ improves the efficiency by a factor $\sim 10^3$ for $N\sim 1.6\times 10^4$, the number of LPs used in our calculations.

We note, however, that this method requires extra computations for virtual-channel interactions. The number of attempts to sample an interaction exceeds the number of actually simulated interactions by a factor $\sim\tilde{\sigma}_{\max}/\tilde{\sigma}_{\text{real}}$. If the behaviour of σ_{real} is singular (e.g. if there is a resonance), a straightforward application of this method would be inefficient. One can avoid this inefficiency by treating a resonant feature as a separate process that occurs only within a narrow energy range.

2.4 Simulating a spatially inhomogeneous system

An inhomogeneous spatial distribution of particles is easily simulated using the LP method. However, the variable particle density must be taken into account when searching for target LPs. In the previous section, we noted that a target LP is searched for among the LPs in the volume surrounding the incident LP. In this volume, the particle distribution is treated as uniform. To reproduce an inhomogeneous distribution, we simply divide the simulation volume into a number of spatial cells. LPs are allowed to interact within the same cell and the mean free time is sampled by using the average densities of target particles inside a cell. Note, however, that in the LP method the trajectories of LPs are followed and the escape or transition to another spatial cell is simulated exactly. Therefore any cell shape can be used, as long as the spatial cell structure is adjusted so that the desired density gradients of target particles are obtained. For time-dependent problems, spatial cells can be adaptive, i.e. their coordinates and shapes can be changed at each time-step.

In this method, the cumulative arrays are pre-computed separately for each cell. This requires little additional memory if properly organized. Discretization of the total simulation volume into spatial cells hardly affects the computing time if the total number of LPs is kept constant.

2.5 Special modifications of the LP method

There are situations in which the standard LP method is inefficient, due to overly frequent interactions or to poor statistics of target LPs. To handle these situations, it was necessary to develop a few specialized techniques. Since these lie somewhat outside the main line of argument in this paper, they are described in the Appendix.

2.6 Comparison of different methods

As Xu (1994) argued, the kinetic equation approach is the most efficient method for treating 0D (i.e. one spatial zone) non-linear problems. To solve the kinetic equations describing the time evolution of the system requires $O(N_p^2N_kN_s)$ calculation steps per iteration (i.e. per time-step), where N_p is the number of bins in momentum space, N_k is the average number of momentum cells within the kinematic limits for an end product of an interaction, and N_s is the number of space cells. For a typical 0D pair cascade calculation (Coppi 1992; Xu 1994), $N_s=1$, $N_p\sim 100$, $N_k\sim 10$ (the latter estimate is very approximate), and the number of calculation steps for one iteration of the whole system is $\sim 10^5$.

The LP method requires $O[N_{\text{LP}}\log(N_{\text{LP}})]$ calculation steps per iteration, where N_{LP} is total number of LPs. In practice, one can neglect the $O(\log N_{\text{LP}})$ factor, as operations associated with this factor are very fast. This estimate does not depend on the number of dimensions of the system and satisfactory statistics are obtained using only $\sim 10^5$ LPs, provided that the system is not too complicated geometrically. Thus for 0D problems the number of steps is of the same order as for the kinetic equation approach. However, each step takes longer to compute using the Monte Carlo method, making the kinetic equation approach more efficient, in agreement with Xu's conclusion.

The situation changes drastically where spatial structure is taken into account. In a 1D problem, for example, the kinetic equation approach requires at least 10 angular bins, making $N_p\gtrsim 10^3$. The required number of spatial cells is at least a few. Thus the number of calculation steps for one iteration is $\sim 10^7$, or even more, since N_k can grow when angular bins are introduced. This is still a reasonable value for simulations done on PC-486 computers. However, the number of steps

required by the LP Monte Carlo method is now two orders of magnitude smaller. Therefore, provided that the LP computation time for one step is not longer than 100 times the kinetic computation time per step, we conclude that the LP method is probably faster already for 1D geometries. Note, however, that it is not easy to compare the computation time per step for the two methods, as interactions are simulated in completely different ways.

We now compare the efficiencies of the two Monte Carlo methods. The PSD approach can be characterized by $O[N_p N_s N_r \log(N_p)]$ steps per iteration, where N_r is the number of times the calculation is repeated for each phase-space cell. The value of N_r is optional, but a larger value will yield better statistics (Xu 1994). Since a simulation in the PSD representation can be interpreted as the evolution of $N_p N_s N_r \log(N_p)$ LPs, and since each LP requires similar computation time in both methods, the relative efficiency of the two methods can be deduced simply by comparing the number $N_p N_s N_r$ with the number of LPs used in the LP simulation. [For reasons given above we neglect the $\log(N_p)$ factor.]

Suppose that the system is represented in both approaches in such a way that $N_{LP} = N_p N_s N_r$, i.e. the computation time in both representations is of the same order. Which approach will then give better statistical and systematic accuracy? The answer depends on the kind of information being sought. To obtain the best statistics in the parts of the particle distributions containing most of the energy, the best strategy is to use an LP method in which all LPs have the same energy, i.e. $E_i(\varepsilon) = \text{constant}$. This automatically introduces a greater number of LPs into the regions of phase space containing most of the energy. If one is interested in the low-luminosity parts of the spectrum, the PSD representation has an advantage due to its use of 'adaptable' statistical weights (Xu 1994), which provides equal statistical accuracy in equal volumes of phase space. The efficiency of simulating the high-luminosity regions is lower, however, since most time is spent simulating the relatively small amount of energy in the low-density regions of phase space. The same goal can be achieved in the LP approach by using an energy-dependent $E_i(\varepsilon)$ (see Section 2.1), but this is less convenient since it requires a priori assumptions about the particle energy distributions.

Another advantage of the PSD approach is that it uses information about the intermediate states of particles which evolve rapidly in energy space due to numerous interactions (see Xu 1994), and therefore gives a better statistical representation of the particle distribution. This is important if the rapidly evolving particles play the role of targets in some process, e.g. electrons in the synchrotron self-absorption process. Simulation of synchrotron self-absorption in the LP approach requires a special treatment, incorporating some features of the PSD representation (see the Appendix).

Summarizing the above, we find that both Monte Carlo methods have comparable efficiencies from a statistical point of view for problems of low dimensionality. The LP representation is more efficient if we are interested in the high-luminosity parts of the spectrum, while the PSD approach is more convenient for reproducing the low-luminosity spectral regions and to simulate processes with very large cross-sections, such as synchrotron self-absorption.

Note, however, that the equation of N_{LP} and $N_p N_s N_r$ is possible only for problems with dimension less than two. For 2D pair-cascade problems the number of required PSD phase-space cells can be estimated to be $> 10^5$, i.e. 100 energy bins times (20×10) angular bins times more than 10 spatial cells. Note that two angles are necessary even in 2D problems. If each cell in phase space is simulated $N_r >$ a few times, the PSD approach is equivalent in computational speed to an LP simulation with a few million LPs. In the LP scheme, on the other hand, $\lesssim 10^5$ LPs is sufficient, except for problems with very complicated geometry. In addition, PSD calculations of higher dimensionality are less efficient because a larger fraction of time is spent on simulating regions of phase space that are sparsely populated. In order to overcome this problem, it is necessary to use complicated adaptable mesh techniques, whereas in the LP representation an optimal distribution of interactions throughout phase space is achieved automatically.

We now comment on Xu's (1994) claim that the LP method is very inefficient for simulations of dynamical (i.e. time-dependent) systems. He argues that the time resolution of the LP method is constrained by memory limitations on the number of LPs, while statistical precision can be achieved in the PSD aproach by performing a large number, N_r, of simulations of each phase-space cell. However, consideration of the memory requirements does not bear this out. Each LP in a 3D problem requires approximately 40 bytes of memory. Thus, for 10^5 LPs (in the present paper, 16 384 LPs were used), 3.5 Mb of memory is needed and computers like a PC-486 can be used. With a few tens of Mb of memory, which Xu claimed is necessary for a 2D problem using the PSD method, one could simulate 10^6 LPs. Noting that the escape rate of LPs is $\sim N_{LP}$ per light crossing time, R/c, and that reasonable statistics are achievable for $\sim 10^5$ LPs, we find that a memory capacity of 10^6 LPs seems sufficient to simulate a system with very good statistical accuracy over a time as short as the light crossing time. Thus we claim that, at present, computational speed rather than memory requirement is the bottle-neck when simulating non-linear radiating systems.

Let us next consider systematic errors. Both approaches have errors associated with the finite value of the time-step, Δt. This does not present a serious problem, however, because these errors $\sim O(\Delta t)$ and can be reduced to any small value by choosing a sufficiently small Δt. The PSD approach involves some additional errors due to the discretization of momentum space as well as real space. The effects of the former errors can be moderate, but effects of the latter can be very unpleasant. A particle in the PSD approach has no continuous trajectory. Motion between space cells is described by Poissonian statistics, i.e. by a constant probability per unit time, $\mathrm{d}p/\mathrm{d}t = p/\tau$, where the escape time τ is defined by the size of the space cell divided by the particle velocity. As a result, the flight times of a particle through a space cell will be distributed with a probability $p \sim \mathrm{e}^{-t/\tau}$, and after crossing many cells the flight times will have a Gaussian distribution. This effect is especially troublesome in dynamical problems. For example, a front of photons emitted by a flare will be artificially dispersed in space. The effect will cause systematic errors even for steady-state problems – e.g. the dispersion of flight times of a photon through some region will be overestimated.

Finally, we emphasize that the space cells that are used in the LP approach do not discretize space. Rather, the LP method introduces discrete levels of the target density, since the targets are assumed to be uniformly distributed within a

given cell. Therefore the requirements regarding the size and shape of space cells can be considerably weaker than for the PSD approach. The main surviving systematic errors are those associated with gradients of energy and angular distributions across a single space cell. Since the locations of all LPs are exactly known, however, it is possible to reduce these errors by calculating the gradients of target densities within each space cell and taking these into account when sampling free path time.

3 SIMULATIONS OF PAIR CASCADES USING THE LP METHOD

3.1 Pair cascades

The simplest model of a non-thermal pair cascade (see, e.g., review by Svensson 1994 and references therein) comprises a uniform region of size R, into which monoenergetic pairs or electrons with Lorentz factor $\gamma_{\max}$ are injected with a luminosity L_e, along with a luminosity L_{UV} of blackbody radiation at temperature T_{UV} (typically in the UV range). Since the electrons lose virtually all of their energy by Compton scattering the soft photons, the entire luminosity L_e is converted into X-rays and gamma-rays. Electron–positron pairs are typically produced in photon–photon interactions between gamma-rays with energies $\varepsilon > 1$ (in $m_e c^2$ units) and X-rays slightly above the pair production threshold $1/\varepsilon$. The radiation from the pairs may produce further generations of pairs, resulting in a pair–photon cascade.

The self-consistent pair and photon distributions, taking into account reprocessing by the pair cascade, depend on only four parameters: (i) the injection compactness $\ell_e = (L_e/R)(\sigma_T/m_e c^3)$; (ii) the soft (or UV) compactness $\ell_{UV} = (L_{UV}/R)(\sigma_T/m_e c^3)$; (iii) the soft blackbody temperature kT_{UV}; and (iv) the injected Lorentz factor $\gamma_{\max}$ of the electrons. There are essentially three regimes for these pair–photon cascades: (i) no pair cascading for $\ell_e <$ a few; (ii) unsaturated pair cascades for a few $\lesssim \ell_e \lesssim 30$; and (iii) saturated pair cascades for $\ell_e > 30$. For saturated pair cascades, essentially every gamma-ray is absorbed in producing an electron–positron pair.

The pairs cool to non-relativistic energies where they thermalize and eventually annihilate at the equilibrium temperature T_e. This temperature is set by the balance between Compton heating and inverse Compton cooling on the self-consistently determined radiation field. The Thomson scattering depth of the cool pairs, τ_T, becomes larger than unity for injection compactness $\gtrsim 20$. Then the interactions between the pairs and the photons will have a noticeable effect on the radiation field, i.e. the UV and X-ray radiation will be Comptonized. The UV blackbody photons will be upscattered to produce a steep soft X-ray power-law spectrum, while the hard X-rays at energies $> 1/\tau_T^2$ will lose energy by scattering on the cool pairs. The latter process will cause a spectral break in which the spectrum steepens towards higher energies. An important diagnostic parameter is the pair yield, PY, which is the fraction of injected power that is converted into pair rest mass. Guilbert, Fabian & Rees (1983) showed that τ_T depends on the pair yield according to $\tau_T \sim (\ell_e PY)^{1/2}$. The pair yield is an increasing function of ℓ_e, but reaches a constant value of about 10 per cent for saturated pair cascades (Svensson 1986).

If the electrons are injected with a power-law distribution in energy, the model acquires two additional parameters, the slope of the injected distribution, Γ, and the minimum injected Lorentz factor, $\gamma_{\min}$. Another two parameters enter if the magnetic field is taken to be non-zero; these can be chosen to be a 'magnetic compactness', $\ell_B \equiv (B^2/8\pi)\sigma_T R/m_e c^2$, and the source size R (or the magnetic field B).

3.2 Main features of the LP cascade code

We have implemented the LP method described in Section 2 in a code aimed mainly at simulating non-linear electromagnetic and hadron–electromagnetic cascades. This code can also be used efficiently to study simpler linear phenomena, e.g. thermal Comptonization. The prototype of the code, presented in Stern (1988), included electromagnetic processes only; later versions have been extended to include processes resulting from injection of relativistic protons (Stern, Svensson & Sikora 1990; Stern, Sikora & Svensson 1992). The types of particles represented by LPs in the code are photons, electrons, positrons, protons, neutrons, and ^{4}He nuclei and their fragments. The interactions taken into account in the present version of the code are Coulomb scattering of electrons and positrons, synchrotron radiation, synchrotron self-absorption, Compton scattering, photon–photon pair production, pair annihilation, pair production in proton–photon interactions, photomeson production (on both protons and neutrons, including p–n charge exchange), inelastic proton–proton interactions, and photodisintegration of ^{4}He and its products (Sikora & Begelman 1992). In the present version of our code, the pions and muons are assumed to be too short-lived to suffer collisions before decaying and, therefore, we treat their decay products as being produced *in situ*. This assumption, however, is correct only for AGNs and, for simulations of radiation processes in galactic X-ray sources, the collisions of charged pions and muons with other particles have to be followed as well.

The microphysics of most electromagnetic processes is represented by the exact expressions of quantum electrodynamics (e.g. Jauch & Rohrlich 1976). The only approximations we use are the following: the ultrarelativistic expressions for synchrotron emission and self-absorption are also applied in the semirelativistic regime, and an approximate expression for the electron–positron Coulomb scattering cross-section is used. The latter assumes that the target particle is non-relativistic and that the energy exchanged is negligible in comparison with the energy of the incident particle. It can be shown that these approximations introduce negligible effects to the results of simulations for typical astrophysical situations. Non-electromagnetic processes, such as photomeson production, photodisintegration and nuclear collisions, presently do not have good theoretical descriptions and are therefore simulated using experimental cross-sections and inelasticities. We do not give references for that data here because the main purpose of this paper is the demonstration of the efficiency of the LP method for simple scenarios involving only electromagnetic processes.

Simulations of all processes, including the spiralling of charged particles in large-scale magnetic fields, are implemented using full 3D kinematics. Charged particles, being coupled to the background plasma through chaotic magnetic

fields, also participate in externally imposed bulk motions. Multiple spatial cells are used to simulate inhomogeneities in the target medium. A few easily reprogrammable subroutines constitute the model-dependent part of the code, while the main part of the code is universal.

The efficiency of the LP code can be characterized as follows. For 2^{14} (= 16 384) LPs the CPU time for simulating the evolution of a pair cascade over the time interval $\sim R/c$, where R is the characteristic size of the system, varies between 0.1 and 0.5 h (depending on the compactness parameter) on computers such as a Sun-4 or a 486DX2-66 PC. Usually, it is necessary to follow the system evolution for ~ 10–$20(R/c)$, in order first to reach a steady state and then to achieve a statistical quality similar to that shown below in the figures.

3.3 Comparison of LP and kinetic equation results

Here we compare the results of our LP cascade code with the kinetic equation results of Coppi (1992). There are earlier works on pair cascades carried out within the framework of the kinetic equation approach (e.g. Fabian et al. 1986; Lightman & Zdziarski 1987; Svensson 1987; Done & Fabian 1989; Ghisellini 1989), but Coppi (1992) contains the most detailed treatment of the microphysics and covers a larger variety of cascade parameters.

Physical space in our calculations is represented by a homogeneous sphere. This corresponds most closely to the one-zone approximation, i.e. a uniform medium with an escape probability, that was used by Coppi (1992). In this series of calculations we follow Coppi in assuming external injection of pairs. The contribution of the injected pairs to the total pair yield is small in all cases except that represented in Fig. 2, where the intense annihilation line is an artefact of the injection assumed by both Coppi and ourselves.

3.3.1 Simple Compton pair cascades

High-energy processes near accreting compact objects often occur in a background of thermal radiation from optically thick matter. Since this thermal radiation emerges in the UV band in AGNs, most previous work on pair-cascade models, including that of Coppi (1992), has included such a UV radiation component. In these models, the main source of pair-producing photons is Comptonization of the UV photons by relativistic electrons.

We compare our results for Compton cascades with those of Coppi, for three cases. Three of the input parameters are kept constant – monoenergetic pair injection with $\gamma_{max} = 10^3$, $kT_{UV} = 10^{-5} m_e c^2$ and $\ell_{UV}/\ell_e = 2.5$ – while ℓ_e takes the values 10, 100 and 1000. The escaping radiation spectra for the three values of ℓ_e are presented in Fig. 1 (solid curves), together with the spectra from fig. 1 in Coppi (1992; dotted curves). Three derived parameters, the total scattering optical depth τ_T, the pair yield and the pair temperature T_e, are compared in Table 1 with the corresponding results from table 1 in Coppi (1992).

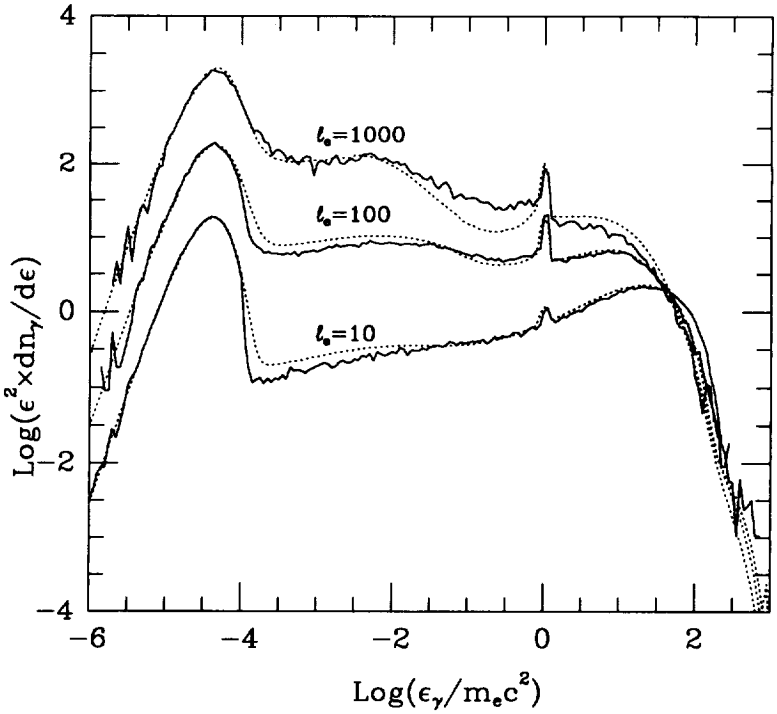

Figure 1. Spectra of escaping radiation from LP calculations (solid lines) and from kinetic equation models (Coppi 1992; dotted lines), for unmagnetized pair cascades with compactnesses $\ell_e = 10$, 100, 1000. $\ell_{UV}/\ell_e = 2.5$ for all cases. The sharper-than-exponential slope of the blackbody spectrum visible in our results is due to our artificially cutting off the exponential tail of the Planck distribution. The blackbody photon energy distribution has been corrected to the full Planck spectrum in calculations presented in Figs 8–12.

Table 1. Comparison of LP and kinetic equation simulations.

$\ell_e = 1000$	τ_T	Pair Yield	kT_e/m_ec^2
LP (this paper)	13.2	0.121	4.5×10^{-4}
kin. (Coppi 1992)	12.8	0.12	6.2×10^{-4}
$\ell_e = 100$			
LP	3.41	0.082	2.3×10^{-3}
kin.	3.34	0.087	2.2×10^{-3}
$\ell_e = 10$			
LP	0.405	0.014	4.9×10^{-3}
kin.	0.502	0.023	5.7×10^{-3}

The differences between the two sets of results, while relatively small, are physically significant. In the regime of unsaturated pair cascades near $\ell_e \sim 10$, the results are known to be sensitive to details of the model, such as how the opacities are calculated and how the radiative transfer is treated. The differences in the radiation spectra for $\ell_e = 10$ may not seem large. A small difference in the X-ray spectrum (and thus the optical depth to pair production), however, causes a small difference in the gamma-ray spectrum, which in turn leads to a substantial difference in the number of pairs produced and thus in the pair yield. Thus Table 1 shows a difference in the pair yield of ~60 per cent.

The main cause for these differences lies in the different treatments of photon escape used in the two methods. Coppi (1992) uses a simple escape probability for his one-zone model. He implements the prescription for the mean escape probability per second, $\dot{P}_{esc}$ (i.e. the inverse of the mean escape time), originally introduced by Lightman & Zdziarski (1987):

$$\dot{P}_{esc} \propto -\dot{n}(\varepsilon)/n(\varepsilon) = [1 + \tau_T f(\varepsilon)]^{-1}, \quad (8)$$

where $f(\varepsilon) = 1$ for $\varepsilon \leq 0.1$, $f(\varepsilon) = (1-\varepsilon)/0.9$ for $0.1 < \varepsilon < 1$, and $f(\varepsilon) = 0$ for $\varepsilon \geq 1$. In the limits of $\tau_T \gg 1$ and $\ll 1$, $\dot{P}_{esc}$ is known analytically for various geometries and spatial distributions of photon sources and scatterers. The choice made by Coppi corresponds to a homogeneous source with slab geometry. For the spherical homogeneous source we consider, $\dot{P}_{esc} = 4c/3R$ for $\tau_T \ll 1$, and $5c/\tau_T R$ for $\tau_T \gg 1$ and $\varepsilon \ll 1$ (Sunyaev & Titarchuk 1980). Fitting the LP simulations, we obtain

$$\dot{P}_{esc} \propto (3/4 + 0.188\tau_T)^{-1} \quad (9)$$

for $\varepsilon \ll 1$, in good agreement with the analytic results.

Coppi's use of equation (8) underestimates $\dot{P}_{esc}$ for a spherical source, resulting in a larger X-ray density which implies, in turn, a larger gamma-ray opacity and the turn-on of pair production at a lower ℓ_e. Thus, the turnover of the photon spectrum for $\ell_e = 10$ in Fig. 1 occurs at a lower energy, and both τ_T and the pair yield (in Table 1) are larger in Coppi's simulation.

For larger values of the compactness, the cascade is saturated. Then τ_T and the pair yield are insensitive to the details and all simulations should give the same results, independent of method. This is clearly seen in Table 1 for $\ell_e = 100$ and 1000. For a fully saturated case, there is effective downscattering and essentially all of the reprocessed gamma-ray power emerges at $\varepsilon \lesssim 1/\tau_T^2$. The radiation spectra obtained using different simulation methods should agree at these energies, simply due to energy conservation. At larger photon energies, photons escape from a surface layer of unit optical depth and there may be large differences in results obtained using different prescriptions for $\dot{P}_{esc}$. Because Coppi underestimates $\dot{P}_{esc}$ for $\varepsilon < 1$, photons spend a longer time in the source, allowing a larger fraction of the photons in this energy range to downscatter to $\varepsilon \sim 1/\tau_T^2$ before escaping. Thus Coppi's spectrum steepens by unity in the spectral index at the hard X-ray break at $\varepsilon \sim 10^{-2}$ for $\ell_e = 10^3$, while our corresponding spectrum steepens by 1/2, as is expected for the case of a uniform distribution of photon sources and scatterers (Sunyaev & Titarchuk 1980). At $\varepsilon > 1$, the opacity due to photon–photon pair production dominates at large ℓ_e. To have too small an X-ray density for $1/\tau_T^2 < \varepsilon < 1$ causes the opacity for gamma-rays to be smaller, and the flux of photons escaping with $1 < \varepsilon < \tau_T^2$ will be too large. This is seen in Fig. 1 for $\ell_e = 10^3$, where Coppi's flux slightly exceeds ours at $\varepsilon > 1$.

The temperature of the cooled pairs given in Table 1 is determined from a balance mainly between Compton cooling and heating, both of which are sensitive to the spectrum. As the spectra obtained with the two methods differ, so will the temperatures. Finally, we remark on an artificial feature visible in Figs 1–6. The sharper-than-exponential slope of the blackbody spectrum is not physical but results from our using a sharp cut-off at the high-energy end of the Planck distribution. This crude approximation is corrected in the models calculated in Sections 3.4.5 and 3.4.6, and in Figs 8–12.

3.3.2 Models including synchrotron radiation

We have modelled the case presented by Coppi (1992) in which the electrons also lose energy through synchrotron emission, with the following parameters: magnetic field $B = 300$ G, magnetic compactness $\ell_B = 2.9$ ($R = 10^{15}$ cm), UV compactness $\ell_{UV} = 6$, UV temperature $kT_{UV} = 1.07 \times 10^{-5} m_e c^2$, power-law injection spectrum with slope $\Gamma = 2.4$, maximum electron energy $= 10^3$, and different minimum injection energies. The agreement is good (Fig. 2), except that the ratio of Comptonization to synchrotron emission is a bit lower in our case, the reason for this being the higher escape probability in our models.

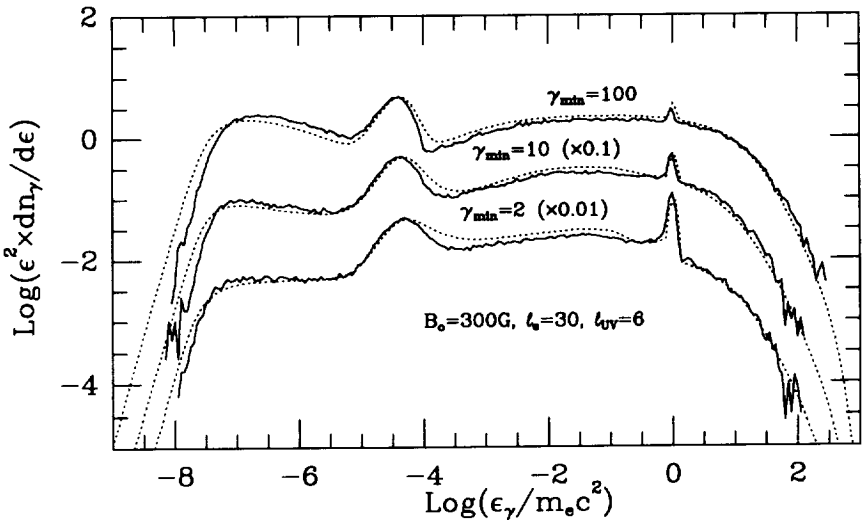

Figure 2. Spectra of escaping radiation from magnetized pair cascades, calculated by the LP method (solid lines) and the kinetic equation method (Coppi 1992; dotted lines). For all models, $B = 300$ G, $\ell_B = 2.9$ ($R = 10^{15}$ cm), $\ell_e = 30$, $\ell_{UV} = 6$, and we assume power-law injection of relativistic electrons with slope $\Gamma = 2.4$, $\gamma_{max} = 1000$, for different values of γ_{min}. Curves are normalized by factors shown in parentheses.

3.4 New results from LP calculations

In this section we present a series of examples illustrating the capabilities of the LP method. We successively model inhomogeneous pair cascades, dynamical effects and complex geometries.

3.4.1 Multishell representation for uniform injection

We start with the same case of a pure Compton cascade and uniform injection as in Section 3.3.1, but with the source volume divided into seven concentric shells, five internal shells where relativistic electrons are injected uniformly and two external shells with zero injection. The internal shell stratification gives a better representation of the spatial distribution of pairs and the spatial and angular distributions of target photons. This allows us to test whether these distributions are roughly uniform inside the source, which might be expected in the case of uniform injection. To emphasize the effect of pair creation in the space surrounding the injection region, we assume a very high electron injection energy, $\gamma_{max} = 10^6$, with $\ell_e = 100$ and $\ell_{UV} = 400$. In Fig. 3 we compare the results with the equivalent one-shell model, in which energy is injected within the same volume as for the multishell model. The difference is visible only in the highest-energy part of the spectrum, where the multishell model gives an exponential cut-off, whereas the one-shell model predicts a power-law tail. The reason for the difference is that the multishell model takes into account pair creation outside of the injection region; due to the very high electron injection energy, gamma rays are effectively absorbed in this region. The X-ray part of the spectrum is insensitive to spatial stratification, which can be explained by the fact that the pair distribution inside the injection region does turn out to be almost uniform. We can therefore conclude that a one-shell approximation for a region of uniform injection is adequate.

3.4.2 Non-uniform injection

We next consider a case in which relativistic electrons and UV radiation are injected non-uniformly. We assume that both injection functions are distributed according to $dL/dR \sim 1/R^2$ and that the ratio of the maximum injection radius to the minimum injection radius is 5. The escaping spectrum is presented in Fig. 4, together with the spectrum for uniform injection (as in Section 3.4.1), both calculated assuming the same multishell structure. The spectra are almost identical in the photon energy range $\varepsilon < 1$, implying that the uniform injection approximation is probably adequate for making predictions about X-ray spectra.

At higher energies, the non-uniform injection model gives a flux that is lower by a factor $\simeq 2$. This can be explained in terms of gamma-ray absorption. In the non-uniform injection case, gamma-rays are weighted towards the central shells. Since the compactness is higher there, the optical thickness for gamma-ray absorption by pair production is also higher and the net flux of escaping gamma-rays is smaller.

3.4.3 Additional heating of thermal pairs

Motivated by results from the OSSE instrument on *Compton Gamma-Ray Observatory* (Johnson et al. 1994), showing that spectra of Seyfert galaxies often have a sharp break at energies corresponding to temperatures much higher than the expected Compton temperatures in pair-cascade models, we calculated a model in which a large fraction of the energy is injected by direct heating of the thermal pairs (i.e. those that have cooled from relativistic energies but have not yet annihilated). Such heating was considered by Zdziarski et al. (1993). It can be caused by Coulomb interactions with semi-relativistic protons or just by an acceleration process of finite efficiency, as was suggested by Zdziarski et al. (1993). High temperatures of scattering electrons can also be mimicked by high velocities of turbulent elements, provided that the

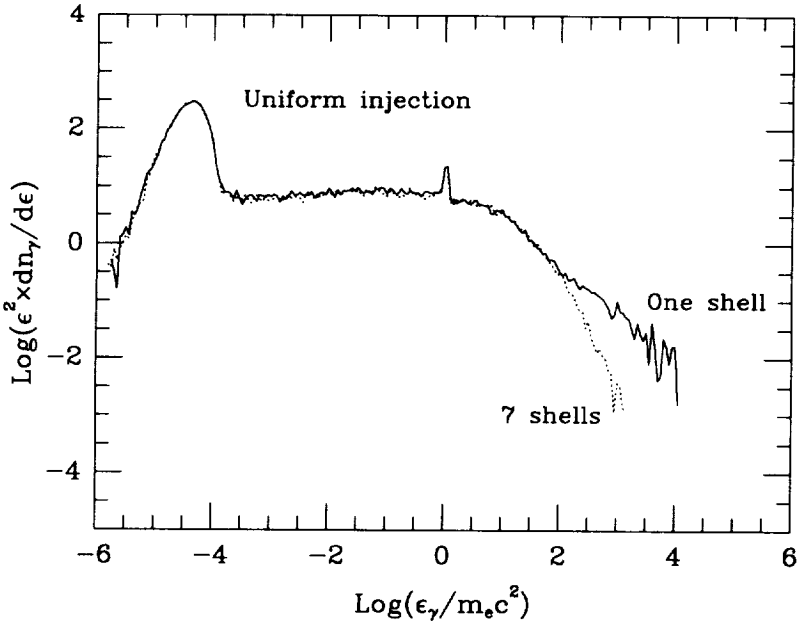

Figure 3. The effects of using multiple spatial shells on pair-cascade spectra calculated using the LP method, for a high electron injection energy, $\gamma_{max} = 10^6$, with $\ell_e = 100$, $\ell_{UV} = 400$. The solid curve is for the one-shell model and the dotted curve is the equivalent model using seven shells.

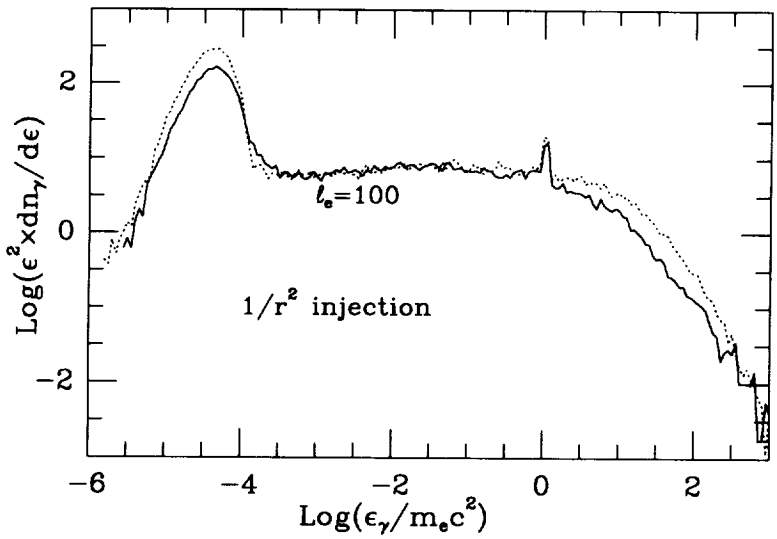

Figure 4. The effects of non-uniform relativistic electron injection on escaping radiation spectra are shown for models with seven spatial shells. Solid curve – electrons are injected according to the radial distribution $dL_{inj}/dr \sim r^{-2}$ within the limits $0.2 < r < 1$. The parameter ℓ_e equals the compactness referred to $r = 1$, and $L_{UV} = 4L_e$. Dotted curve – model with the same parameters, but with uniform electron injection.

characteristic size of turbulent cells does not exceed the photon mean free path.

In our calculations we artificially set a constant temperature and performed simulations for the same shell structure as in the previous two examples. The power that must be supplied to the pairs to support the assumed temperature is found to be $\simeq 3$ times higher than the power injected through relativistic electrons. The results are presented in Fig. 5 and compared there with the purely thermal model.

The lack of a UV bump means that practically all UV photons undergo Comptonization. In order to reproduce observed AGN spectra, which generally possess prominent UV bumps, one would have to appeal to additional geometric complexity in the source. For example, one could posit that the UV source is spatially separated from the cascade region, or that the active corona only partially covers the accretion disc. Partial covering could result from the local character of coronal activity (magnetic flares), and the domination of

thermal over non-thermal Comptonization could occur if most of the energy dissipated in flares goes to heat the plasma.

3.4.4 Effect of radial bulk motion

To study the possible effects of a radial bulk motion on the spectrum, we made calculations for the same set of parameters as in Section 3.4.3, adding a constant inward radial velocity (accretion) and assuming an r^{-2} distribution of electron injection. In this case the temperature is fixed in the rest frame of the pair flow. Escaping radiation spectra for $v/c = -0.1$ are shown in Fig. 6 in comparison with the same case for $v = 0$.

As expected, accretion hardens the spectrum. This effect is essentially a form of adiabatic heating, in the sense that the compressional force does work against the photon gas. In terms of radiation transport it can be described as first-order Fermi acceleration of photons in the converging flow: a photon is more likely to be scattered by an oncoming electron than by a receding one, leading to a mean increase of photon energy per scattering. The mean number of scat-

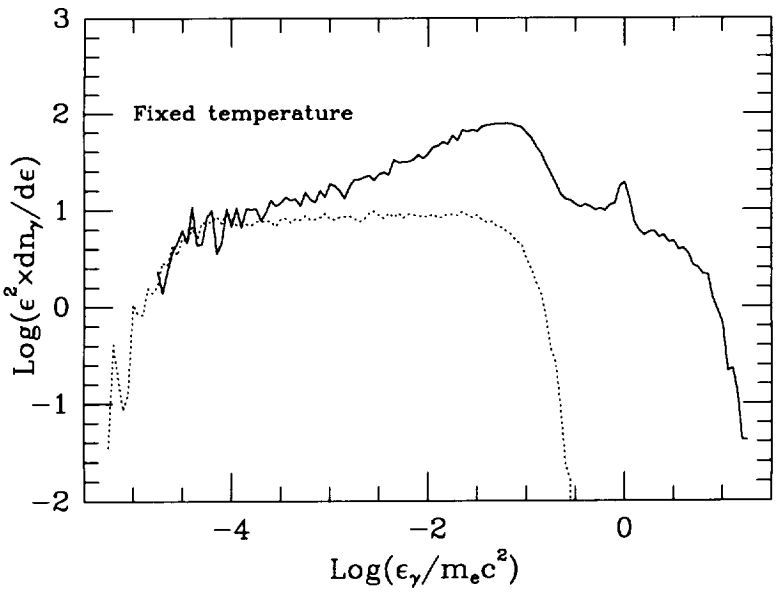

Figure 5. The effects of additional pair heating. The temperature of the thermal pair distribution was artificially supported at the constant value $kT = 0.03 m_e c^2$. Solid curve – Comptonized cascade spectrum, for injected power with compactness parameters $\ell_e = 100$, $\ell_{UV} = 10$. The total escaping luminosity L_{esc} is three times larger than the injected power, with the additional energy supplied through pair heating. Dotted curve – Comptonization of the UV Planck spectrum by heated thermal pairs alone, for $\ell_e = 0$, $\ell_{UV} = 10$ and $kT = 0.03 m_e c^2$. The density of thermal pairs was set equal to that obtained in the non-thermal cascade calculation.

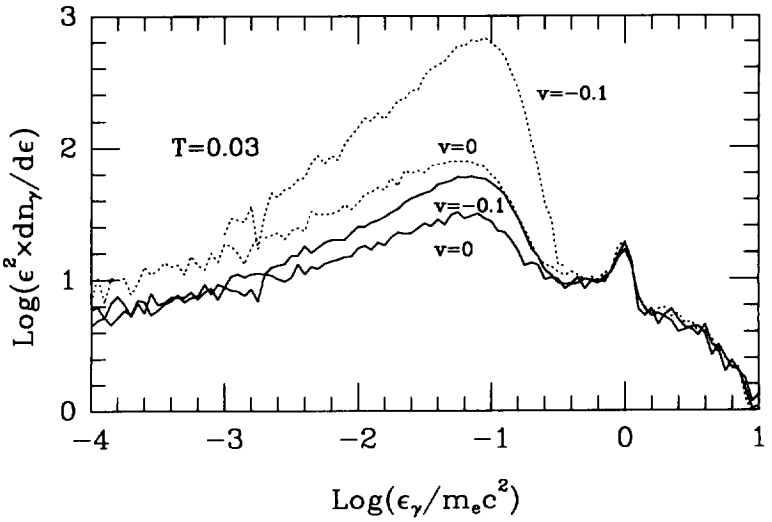

Figure 6. The effect of radial bulk motion with constant inflow velocity $v = -0.1c$ on the escaping spectrum at fixed pair temperature $kT = 0.03 m_e c^2$, for $\ell_e = 100$, $\ell_{UV} = 10$ and an r^{-2} radial distribution of electron injection. Dotted curves – no photon absorption in the centre is simulated. Solid curves – photons are absorbed at $r < 0.15$.

terings becomes 26.4, versus 11.3 for $v=0$, but the optical depth is almost the same as the model with $v=0$. The escaping luminosity is 17.6 times larger than the non-thermal power, and a thermal break in the escaping spectrum is almost transformed in shape into a Wien peak.

In the case shown by dotted lines in Fig. 6, all photons are allowed to escape. In reality, a large fraction of the photons produced in the central region of an accretion flow could be absorbed by the black hole. This case is approximated by our second model (solid lines in Fig. 6), in which all photons and pairs that venture within $0.15R$ of the centre are killed, where R is the radius of the region where energy injection is taking place. For the case of accretion with $v=-0.1$, the effect is very strong – the escaping luminosity drops almost by an order of magnitude and the mean number of scatterings before escape drops by more than a factor of 2. For $v=0$, the effect of absorption at the centre is smaller but still significant.

3.4.5 Two-dimensional simulation of a non-thermal corona above an accretion disc

Reflection and reprocessing of X-ray radiation by optically thick gas (e.g. an accretion disc) is thought to be an important feedback mechanism in the production of high-energy spectra in AGNs (Haardt & Maraschi 1993). All geometrical configurations that include a disc, except perhaps an infinite slab geometry, are at least two-dimensional. We calculated two simple models to illustrate how the LP method could handle both the higher dimensionality and the feedback self-consistently. As in previous models we assume that the thermal particles in the corona are subject to an additional heat source. Instead of fixing the temperature as in the previous two subsections, however, we fix the power supplied to the thermal electrons L_t and calculate the temperature self-consistently, as well as the Compton reflection component from the disc and the final escaping spectrum.

As a simplified model for reprocessing, we follow each photon entering the disc, taking into account only Compton scattering (photoelectric absorption is neglected in these calculations, but is included in the calculations presented in Section 3.4.6). If the photon is reflected after fewer than 10 Compton scatterings, we follow its further trajectory, starting with the energy at which it leaves the disc. The energy deposited in the disc is emitted in the form of a blackbody UV photon from the point at which the incident photon entered the disc. If a photon remains in the disc through 10 or more Compton scatterings, we assume that the whole photon energy is re-emitted in the form of UV blackbody radiation with a specified temperature.

We consider two geometrical configurations: (a) hemisphere + disc; and (b) finite cylinder + disc.

(a) *Hemisphere + disc.* This model consists of an infinite plane parallel slab ('disc') surmounted by a finite hemispherical region of hot plasma above the disc surface (Fig. 7a). The reflection and re-emission from the disc beyond the hemisphere is taken into account. This configuration can be associated with two physical scenarios – a global hot plasma corona (in which case the black hole must be located in the centre of the hemisphere) and a local bulge of hot plasma on the surface of the disc. The spatial cell configuration is shown in Fig. 7.

We performed two calculations for this geometry. The first is a test of pure thermal Comptonization, with reprocessing and reflection. The hemisphere was homogeneously filled by electrons with an optical depth $\tau_T=2$ across the radial distance. We set the ratio $L_t/L_{UV}=25$, so that the intrinsic blackbody emission of the disc is negligible and the main share of UV radiation is produced as the result of the feedback. The absolute value of the power does not matter as we have no photon–photon interactions and the result can be scaled to any total power as well as to any size of the source. The resulting spectrum in direction range $\cos\theta>0.5$ (measured with respect to the disc normal) is shown in Fig. 8

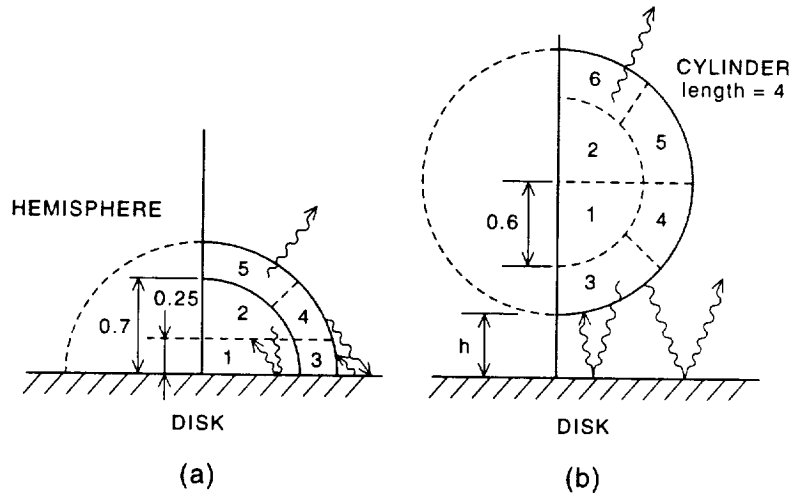

Figure 7. Geometric arrangements of spatial cells used to model hot plasma + disc systems. The disc is modelled by an infinite plane slab; numbers mark the spatial cells in the regions of hot plasma. All sizes are given in terms of the radius of the hot plasma region, $r=1$. (a) Hot plasma occupies a hemisphere sitting on the disc surface. The coordinates of the LPs are two-dimensional. (b) Hot plasma occupies a cylinder of length 4, the lower edge of which is situated a distance h above the disc surface. In this case the coordinates are three-dimensional, but simulations are performed for only one-half of the cylinder. Full dimensionality is achieved by mirror reflection of all photons crossing the median plane.

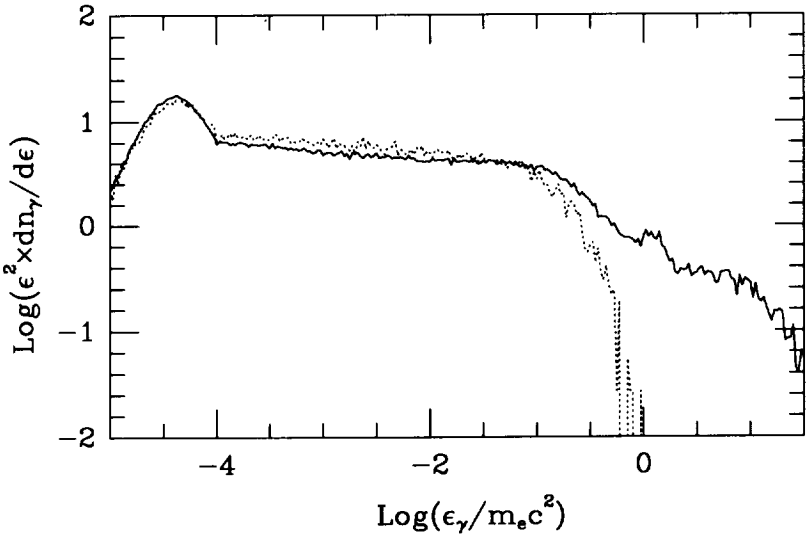

Figure 8. Escaping photon spectra for the case of hemisphere + disc in the angular range $\cos\theta > 0.5$, where θ is the angle from the disc normal (see text and Fig. 7). Solid curve – combined non-thermal and thermal power supplies: $\ell_e = 10$, $\ell_t = 50$. Dotted curve – only thermal power supply to ambient electrons: $\ell_t = 50$, $\tau_T = 2$.

(dotted line). Its slope corresponds to an energy spectral index $\alpha \approx 1.09$. Note that in this geometry it is impossible to get much harder ($\alpha < 1$) spectra due to strong feedback, unless $\tau_T < 1$ (Haardt & Maraschi 1993). The electron temperature in the hemisphere is found to vary as a function of position, increasing from $kT = 0.06 m_e c^2$ at the bottom to $0.14 m_e c^2$ at the top.

In the second model we do not assume the existence of any ambient thermal plasma. Instead, only relativistic electrons are injected. The injection parameters in this model are $\ell_e = 10$, $\ell_t = 50$, $\ell_{UV} = 2$ and $\gamma_e = 10^6$. The resulting spectrum is shown in Fig. 8 (solid line). Here the spectral index is $\alpha = 1.06$, and the temperature of the thermal pairs in the hemisphere increases from 0.10 at the bottom to 0.17 at the top. The pair density distribution is close to uniform, with a maximum density contrast (ratio of the maximum density to the minimum density) of 1.34.

The anisotropy of the radiation can be estimated from Fig. 9, which shows the spectrum of the latter model for two ranges of polar angle, $\cos\theta > 0.5$ (solid line) and $\cos\theta < 0.5$ (dotted line).

(b) *Cylinder (magnetic flux tube) + disc.* The feedback can be weaker if the region of hot plasma is separated by some distance from the surface of the disc. We represent this kind of geometry, which could be associated with a magnetic flux tube in the corona, by a cylinder of radius r whose axis lies parallel to the disc surface. The length of the cylinder is $4r$, and the gap h between the cylinder and the disc is $0.5r$. The geometry of the spatial cells is shown in Fig. 7(b). The cylinder was not divided into space cells along its axis, implying a 2D target density distribution within the cylinder, but particle escape was treated fully three-dimensionally.

As above, we assumed $L_t/L_{UV} = 25$, where L_{UV} is the intrinsic disc radiation penetrating into the cylinder. We calculate the same variants as in the previous geometry. For the case of pure thermal Comptonization when electrons fill the cylinder homogeneously with $\tau = 2$ (from the axis to the surface), the resulting spectral index is $\alpha = 0.83$ (Fig. 10, dotted line) and the temperature increases from the side closest to the disc to the side furthest from the disc, from 0.06 to 0.10. The UV bump is dominated by the reprocessing of hot plasma radiation by the disc.

In the second case we use injection parameters $\ell_e = 10$ and $\ell_t = 50$ (obtained by dividing the appropriate luminosity by the radius of the cylinder, so that the total power and effective compactness are higher than in the case of the hemisphere). The resulting spectral index is $\alpha = 0.86$. Again, temperature increases from the bottom to the top, from 0.08 to 0.15. The pair distribution is almost uniform, with a maximum density contrast of 1.24.

The calculated feedback coefficient for this geometry is $L_{fb}/L_{tot} = 0.14$, where L_{fb} is the luminosity re-emitted by the disc and entering the hot plasma region, and L_{tot} is the total power supplied to the hot plasma. This feedback coefficient is probably the main parameter defining the maximum X-ray slope that can be achieved in a given geometry with multiple thermal Comptonization. For example, for the case of a spectrum with $\alpha < 0.8$ and with a thermal break such as that of NGC 4151, we can conclude that the separation of emitting region from the disc is higher than in the case considered here. To demonstrate the effect of spatial separation between the region of hot plasma and the disc, we simulated a hybrid model with the same parameters as in the previous example, except for a larger gap $h = 1.5r$ between the cylinder and the disc. The result is presented in Fig. 11 (dotted line), in comparison with that for $h = 0.5r$ (solid line). The spectral index for the case $h = 1.5r$ is $\alpha = 0.7$, and the temperature and pair density are ~ 10 per cent higher than in the case $h = 0.5r$.

3.4.6 Thermal pair production and effect of photoelectric absorption in the disc

Thermal pair production is one of the effects which can regulate the optical depth of the hot plasma (Ghisellini &

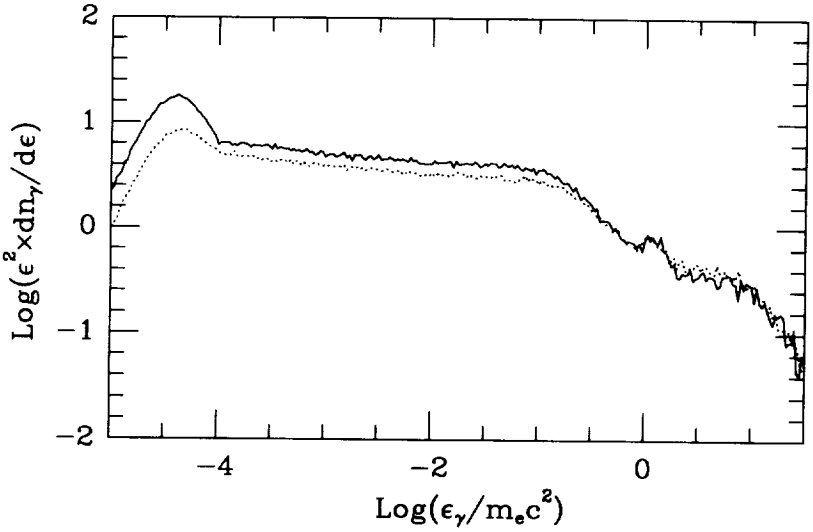

Figure 9. Escaping photon spectra for the case of hemisphere + disc, combined non-thermal and thermal power supplies as in Fig. 8, for two angular ranges. Solid curve – cos $\theta > 0.5$. Dotted curve – cos $\theta < 0.5$.

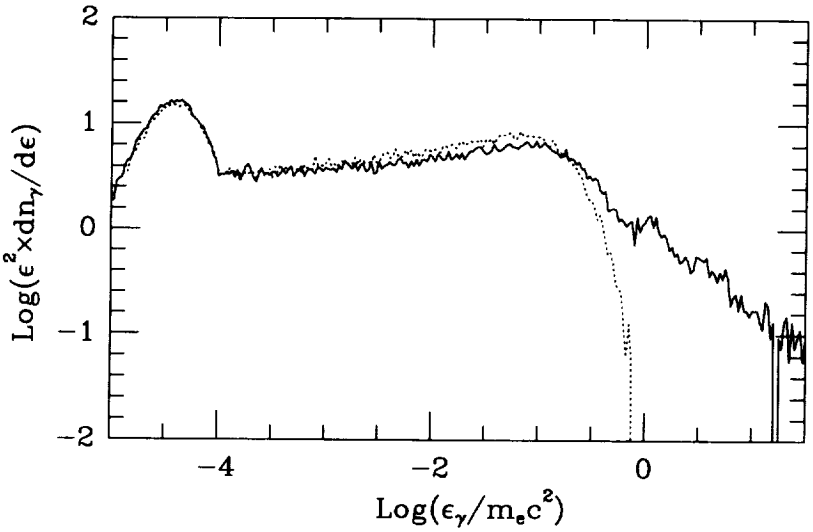

Figure 10. Escaping photon spectra for the case of cylinder + disc with $h = 0.5$, in the angular range cos $\theta > 0.5$ (see text and Fig. 7). Solid curve – combined non-thermal and thermal power supplies: $\ell_c = 10$, $\ell_t = 40$. Dotted curve – only thermal power supply to ambient electrons: $\ell_t = 50$, $\tau_t = 2$ over the distance of the cylinder's radius.

Haardt 1994, and references therein). We demonstrate the effects of thermal pair production for the same geometry (cylinder + disc, $h = 0.5r$) as in the previous section. The thermal power supply corresponds to $\ell_t = 50$. This is the same as the total power in our hybrid model, the difference being that the non-thermal component is now absent. No pre-existing matter has been assumed to occupy the cylinder, so that the entire optical depth of the hot plasma is the result of pair production. We have also included photoelectric absorption in the disc, to demonstrate the formation of the 'reflection hump' (Guilbert & Rees 1988; Lightman & White 1988). (In our previous calculations neglecting photoabsorption, the reflected spectra have a shape close to the shape of the incident spectrum at energies $\varepsilon < 0.1$.) To calculate the photoabsorption cross-section we used the code of Wilms (1994, private communication), based on data from Verner et al. (1993), assuming cosmic abundances (Grevesse & Anders 1989; Shull 1993 for Fe and Cl) and zero ionization stage. The resulting average optical depth along the radius of the cylinder is 0.30, with the maximum gradient of the plasma density $d\tau/dr$ ranging from 0.21 in the region closest to the disc to 0.37 in the inner upper part of the cylinder (Fig. 7a, cell 2). The temperature ranges from 0.32 in the inner lower part (Fig. 7a, cell 1) to 0.44 near the upper surface of the cylinder. The resulting spectra for two angular ranges are shown in Fig. 12. The dominating slope is $\alpha = 0.81$. The feature near $\varepsilon = 0.01$ is a combination of the fluorescent iron Kα line at 6.4 keV and the iron K edge, which are un-

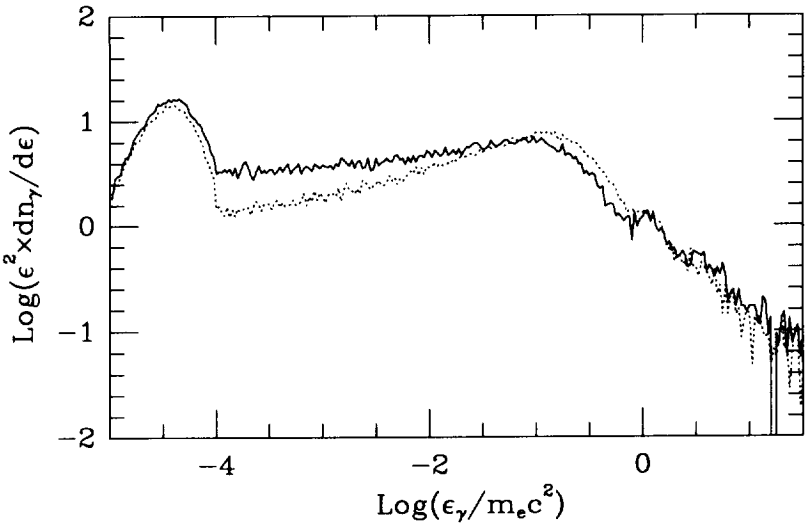

Figure 11. Escaping photon spectra for the case of cylinder + disc, showing effects of changing the distance of the cylinder above the disc. Solid curve – $h = 0.5$; same model as in Fig. 10. Dotted curve – same parameters, except that $h = 1.5$.

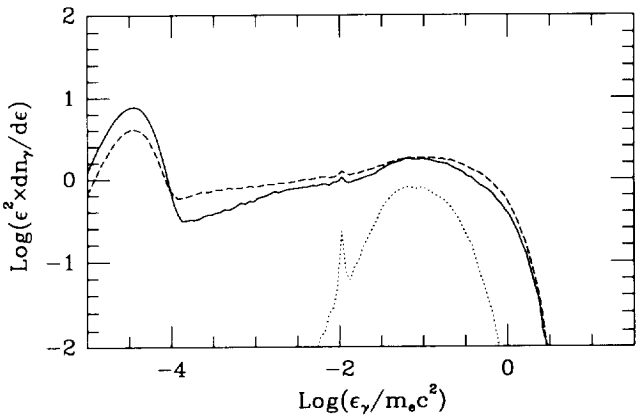

Figure 12. Escaping photon spectra for the case of cylinder + disc and pure thermal pair production with detailed simulation of photon interactions in the disc. The geometry ($h = 0.5$) and total power ($\ell_t = 50$) are the same as in Fig. 10. Solid curve – total escaping spectrum in angular range $\cos\theta > 0.5$. Dashed curve – $\cos\theta < 0.5$. Dotted curve – only the reflected component for $\cos\theta > 0.5$.

resolved. The deviation of the 'face-on' spectrum (solid line) from a power law in the soft X-ray range is probably associated with a lower contribution due to single upscattering when the viewing angle coincides with the main direction of the UV flux. The reflection features as well as the UV bump are stronger in the face-on spectrum because the intensities of both are proportional to $\cos\theta$.

4 CONCLUSIONS

The comparison of our results with those of the kinetic equation approach demonstrates an agreement good enough to conclude that it provides a successful test of the LP code. The agreement is better than one might expect, considering that the different treatments of radiative transport lead to significant differences in escape probabilities. When we introduce a more realistic treatment of radiation transport into the LP code through the multishell representation, the essential differences appear mainly in the hard ($\varepsilon_\gamma > 1$) part of the escaping luminosity spectrum.

We therefore conclude that, in the case of a simple electromagnetic cascade with a modest degree of Comptonization and moderate compactness, the escaping spectrum in the range $\varepsilon_\gamma < 1$ is not very sensitive to details of radiation transport. The kinetic equation approach as well as the PSD Monte Carlo scheme can be successfully applied in this case, especially if one chooses the correct escape probability that follows from our results or can be estimated from simplified linear Monte Carlo simulations of photon transport.

The advantages of the LP method over other approaches increase rapidly with the geometric complexity of the system being modelled. For problems such as the accretion disc plus corona considered in Section 3.4.5, emission from relativistic jets, etc., the correct 2D or 3D treatment of the radiative transport becomes crucial. We believe that such problems can be solved with the PSD representation, but only at the expense of larger computer resources and with considerably more effort put into tailoring the code to the specific model. The calculations presented in this paper have all been made on Sun-4 and PS-486 computers, with a CPU time for each run of at most a few hours. In view of the statistical precision obtained in these calculations (as estimated from the fluctuations of the spectral curves shown in the figures), we feel that the LP method holds considerable promise for simulating non-linear radiating systems in astrophysics.

ACKNOWLEDGMENTS

This work was supported in part by NASA grants NAG 5-2026 (*CGRO* Guest Investigator Program) and NAGW-766 (Theoretical Astrophysics Program), NSF grants AST 91-20599 and INT 90-17207, and the Polish State Committee for Scientific Research grant 221129102.

REFERENCES

Chen K., Ruderman M., 1993, ApJ, 402, 264
Coppi P. S., 1992, MNRAS, 258, 657
Coppi P. S., Blandford R. D., Rees M. J., 1993, MNRAS, 262, 603
Daugherty J. K., Harding A. K., 1982, ApJ, 252, 337
Done C., Fabian A. C., 1989, MNRAS, 240, 81
Fabian A. C., Blandford R. D., Guilbert P. W., Phinney E. S., Cuellar L., 1986, MNRAS, 221, 931
Ghisellini G., 1989, MNRAS, 238, 449
Ghisellini G., Haardt F., 1994, ApJ, 429, L53
Grevesse N., Anders E., 1989, in Waddington C. J., ed., Cosmic Abundances of Matter. AIP Conf. Proc., 183, 1
Guilbert P. J., Rees M. J., 1988, MNRAS, 233, 475
Guilbert P. J., Fabian A. C., Rees M. J., 1983, MNRAS, 205, 593
Haardt F., Maraschi L., 1993, ApJ, 413, 507
Jauch J. M., Rohrlich F., 1976, The Theory of Photons and Electrons, 2nd edn. Springer-Verlag, Berlin
Johnson W. N. et al., 1994, in Proc. Second Compton Symp. AIP, New York, in press
Kirk J. G., Mastichiadis A., 1989, A&A, 211, 75
Lightman A. P., White T. R., 1988, ApJ, 335, 57
Lightman A. P., Zdziarski A. A., 1987, ApJ, 319, 643
Lingenfelter R. E., Hua X., 1991, ApJ, 381, 426
Madejski G. M. et al., 1994, ApJ, in press
Mészáros P., Rees M. J., 1993, ApJ, 418, L59
Pozdnyakov L. A., Sobol I. M., Sunyaev R. A., 1977, SvA, 21, 708
Rees M. J., Mészáros P., 1992, MNRAS, 258, 41P
Shull J. M., 1993, Phys. Scr., T47, 165
Sikora M., 1994, ApJS, 90, 923
Sikora M., Begelman M. C., 1992, in Stenger V. J., Learned J. G., Pakvasa S., Tata X., eds, High Energy Neutrino Astrophysics. World Scientific, Singapore, p. 114
Stern B. E., 1985, SvA, 29, 306
Stern B. E., 1988, Nordita/88-51 A, preprint
Stern B. E., Svensson R., Sikora M., 1990, in Miller H. R., Wiita P. J., eds, Variability of Active Galactic Nuclei. Cambridge Univ. Press, Cambridge, p. 229
Stern B. E., Sikora M., Svensson R., 1992, in Holt S. S., Neff S., Urry C. M., eds, Testing the AGN Paradigm. AIP, New York, p. 313
Sunyaev R. A., Titarchuk L. G., 1980, A&A, 86, 121
Sunyaev R. A. et al., 1992, ApJ, 389, L75
Svensson R., 1986, in Mihalas D., Winkler K.-H., eds, Proc. IAU Colloq. 89, Radiation Hydrodynamics in Stars and Compact Objects. Springer-Verlag, Berlin, p. 325
Svensson R., 1987, MNRAS, 227, 403
Svensson R., 1994, ApJS, 92, 585
Verner D. A. et al., 1993, Atomic Data and Nuclear Data Tables, 55, 233
Xu Y., 1994, Internal JILA Report
Zdziarski A. A., Lightman A. P., Maciołek-Niedźwiecki A., 1993, ApJ, 414, L93

APPENDIX A: SPECIAL MODIFICATIONS OF THE LP METHOD

A 1 Accumulation

In many problems, some interactions are very soft (i.e. the energy transfer $\Delta\varepsilon$ is small) and occur too frequently to permit one to follow the evolution of an LP by taking into account each interaction. To treat such interactions we use an *accumulation* technique, in which a large number (F) of interactions is 'accumulated' into a single interaction. The energy transferred in a cumulative interaction is $F\Delta\varepsilon$ and the cross-section of this interaction is $\sigma_{\rm acc} = \sigma/F$. The factor F can be chosen so that the fractional energy change, f, of the incident particle is small but not negligible, say $f = F\Delta\varepsilon/\varepsilon \sim 0.1$.

As an example, consider Compton scattering of soft photons on ultrarelativistic electrons. The change in electron energy (in $m_e c^2$ units) is $\Delta\varepsilon_e \sim \varepsilon_\gamma \varepsilon_e^2 \ll \varepsilon_e$, so that $F \sim f/(\varepsilon_\gamma \varepsilon_e)$ and $\sigma_{\rm acc} \sim \sigma_T \varepsilon_\gamma \varepsilon_e / f$, where σ_T is the Thomson cross-section. Note that $\sigma_{\rm acc}$ is proportional to ε_γ, making the interaction rate proportional to the photon energy density.

A 2 Time-averaged target

There exist situations in which LP statistics are not sufficient to simulate a process accurately. This happens when the cross-section of a process is very large and the density of target particles is small. Examples of such processes are induced Compton scattering and synchrotron self-absorption. In this case, the LP technique can be modified by incorporating some features of the PSD representation.

Let us illustrate the method using the example of synchrotron self-absorption. When simulating this process, hard electron LPs ($\gamma_e \sim 10$–10^2) interact with soft photons. The number of hard electron LPs at any moment is too small to provide satisfactory statistics, because a hard electron loses energy very rapidly. This feature can, however, be turned to our advantage, if we store the intermediate steps in the evolution of a hard electron LP and use the results to construct a time-averaged energy distribution for the electron LPs. Then the soft photon LP can be simulated as interacting with the time-averaged distribution.

This method has the same disadvantages as the PSD representation – for an inhomogeneous, anisotropic system, a multidimensional phase-space density array averaged over time is required.

A 3 Electron thermal pool

The detailed simulation of non-relativistic quasithermal electrons is very time-consuming, because their energy changes resemble a random walk in energy space at a very high speed. Fortunately, we know that the result of this random walk will be very close to a Maxwellian distribution, since the Coulomb coupling among low-energy electrons is generally rapid enough for thermalization. This allows us to use a simplified method to treat the non-relativistic electrons as a 'thermal pool'.

The following is a summary of how we treat this component.

(1) We introduce an electron pool temperature, T_e, and an upper cut-off energy, $E_c > kT_e$. At each time-step, the energies of all electrons in the pool are randomly sampled from a Maxwellian distribution with a cut-off energy E_c.

(2) Each electron crossing E_c by losing energy enters the pool, while an electron crossing E_c by gaining energy leaves the pool. All individual interactions of LPs in the pool are turned off, with the exception of pair annihilation.

(3) The change in the pool energy ΔE_p is calculated at each time-step. This includes the energy transferred in Compton scattering and Coulomb interactions with particles outside the pool and the kinetic energy of electrons leaving and entering the pool. The temperature of the pool is adjusted accordingly.

There is a problem associated with measuring the temperature of the pool. If we defined the temperature directly using energy balance and total statistical weight of pool electrons, we would get statistical fluctuations of the temperature (even including negative values) that were too large. The reason is that the energy flux through the thermal pool (heated by hard photons and electrons through Compton and Coulomb scattering, respectively, and cooled by Compton scattering on soft photons) is large and the total kinetic energy in the pool is comparatively small. To avoid large fluctuations we introduce an artificial relaxation time which forces the temperature to adjust gradually.