
Fluid, kinetic and hybrid approaches for neutral and trace ion edge transport modelling in fusion devices

Published in:
Nuclear Fusion

DOI:
10.1088/1741-4326/ac3fe8

Published: 01/08/2022

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
Fluid, kinetic and hybrid approaches for neutral and trace ion edge transport modelling in fusion devices

To cite this article: D.V. Borodin et al 2022 Nucl. Fusion 62 086051

View the article online for updates and enhancements.

You may also like

- OSCILLATOR STRENGTHS FOR NEUTRAL TECHNETIUM
  R. H. Garstang

- Modeling Coronal Mass Ejections with the Multi-Scale Fluid-Kinetic Simulation Suite
  N. V. Pogorelov, S. N. Borovikov, I. A. Kryukov et al.

- Top down approach to 6D SCFTs
  Jonathan J Heckman and Tom Rudelius
Fluid, kinetic and hybrid approaches for neutral and trace ion edge transport modelling in fusion devices

D.V. Borodin¹, F. Schluck¹, S. Wiesen¹, D. Harting¹, P. Börner¹, S. Brezinsek¹, W. Dekeyser², S. Carli², M. Blommaert², W. Van Uytven², M. Baelmans², B. Mortier³, G. Samaey³, Y. Marandet⁴, P. Genesio⁴, H. Bufferand⁵, E. Westerhof⁶, J. Gonzalez⁸, M. Groth⁷, A. Holm⁷, N. Horsten⁷ and H.J. Leggate⁸

¹ Forschungszentrum Jülich GmbH, Institut für Energie- und Klimaforschung – Plasmaphysik, Partner of the Tri lateral Euregio Cluster (TEC), 52425 Jülich, Germany
² KU Leuven, Department of Mechanical Engineering, Celestijnenlaan 300, 3001 Leuven, Belgium
³ KU Leuven, Department of Computer Science, Celestijnenlaan 200A, 3001 Leuven, Belgium
⁴ Aix-Marseille Univ., CNRS, PIIM, F-13013 Marseille, France
⁵ CEA, IRFM, F-13108 Saint-Paul-lez-Durance, France
⁶ DIFFER—Dutch Institute for Fundamental Energy Research, Partner in the Tri lateral Euregio Cluster, De Zaale 20, 5612 AJ Eindhoven, Netherlands
⁷ Department of Applied Physics, Aalto University, Espoo, Finland
⁸ National Centre for Plasma Science and Technology, Dublin City University, Glasnevin Dublin 9, Ireland

E-mail: d.borodin@fz-juelich.de

Received 16 July 2021, revised 24 November 2021
Accepted for publication 3 December 2021
Published 7 July 2022

Abstract

Neutral gas physics and neutral interactions with the plasma are key aspects of edge plasma and divertor physics in a fusion reactor including the detachment phenomenon often seen as key to dealing with the power exhaust challenges. A full physics description of the neutral gas dynamics requires a 6D kinetic approach, potentially time dependent, where the details of the wall geometry play a substantial role, to the extent that, e.g., the subdivertor region has to be included. The Monte Carlo (MC) approach used for about 30 years in EIRENE (Reiter et al 2005 Fusion Sci. Technol. 47 172–86), is well suited to solve these types of complex problems. Indeed, the MC approach allows simulating the 6D kinetic equation without having to store the velocity distribution on a 6D grid, at the cost of introducing statistical noise. MC also provides very good flexibility in terms of geometry and atomic and molecular (A & M) processes. However, it becomes computationally extremely demanding in high-collisional regions (HCRs) as anticipated in ITER and DEMO. Parallelization on particles helps reducing the simulation wall clock time, but to provide speed-up in situations where single trajectories potentially involve a very large number of A & M events, it is important to derive a hierarchy of models in terms of accuracy and to clearly identify for what type of physics issues they provide reliable answers. It was demonstrated that advanced fluid neutral models are very accurate in HCRs, and at least an order of magnitude faster than fully kinetic simulations. Based on these fluid models, three hybrid fluid–kinetic approaches are introduced: a spatially
Introduction

Numerical simulations with a kinetic code such as EIRENE [1] are indispensable for both understanding and predicting the fuel and impurity transport in the edge and divertor areas of fusion devices. The transport determines impurity penetration towards the core, plasma exhaust and plasma–surface interaction (PSI) issues having impact on the duty cycle of ITER and DEMO. The insight into the interplay of transport and atomic–molecular (A & M) processes [2, 3] provided by modelling is key for understanding of the detachment phenomenon [4, 5], which is critical for many exhaust regimes envisaged for ITER and DEMO.

EIRENE is a multi-purpose Boltzmann-equation Monte Carlo (MC) solver typically employed in an iterative scheme with a fluid plasma code, succinctly referred to as computational fluid dynamics (CFD) code in this paper, see figure 1. A number of 2D CFD-EIRENE code packages such as SOLPS-ITER [6, 7], EDGE2D-EIRENE [8], and SOLEdge2D-EIRENE [9] are extensively employed by the fusion community. Also 3D codes are applied, such as EMC3-EIRENE [10] or SoLEdge3X-EIRENE [11], and are more demanding in terms of CPU and memory requirements. All the mentioned kinetic–fluid packages are under active development. They provide self-consistently generated plasma distributions of macroscopic properties (2D or 3D), heat and particle fluxes to the wall, synthetic spectroscopy and radiative losses. An essential part of the neutral MC tracing procedure are the collisional-radiative models (CRMs) for A & M processes involving main-plasma species and intrinsic/extrinsic impurities. This includes ionization–dissociation–recombination of A & M species, break-up chains of molecules in plasma, and elastic processes.

The plasma fluid side of the packages is typically much less CPU time demanding than the EIRENE side. However, due to the strong non-linear coupling between neutrals and plasma, a large number of iterations of the coupled code systems is needed to reach a statistically stationary steady state. Despite hundreds or even thousands of CFD-EIRENE iterations are typically necessary to converge to a steady state solution, all the established packages mentioned above mostly provide sufficient performance allowing transport calculations for current devices. The situation becomes more difficult for ITER [12], DEMO and other large devices, for which the equilibration of the balance equations becomes much slower. In particular the particle balance poses a bottleneck, because the time scale for the particle balance \( \tau^p \) scales as \( 1/(1 – R) \) [13], with the effective particle recycling coefficient \( R \) approaching unity for ITER and DEMO configurations. The time constant \( \tau^p \) may be in the range of seconds for ITER and DEMO, while time steps in the simulations are typically restricted to the order of \( 10^{-6} \) s or lower due to numerical stability constraints including stiffness on the CFD side of the package. Energy and momentum have much lower effective recycling coefficients, leading to much faster equilibration and posing less of a constraint. Also important is the level of approximation, for instance turbulent transport is often modelled using the gradient diffusion approximation, because full turbulence simulations are much more computationally demanding [11]. In fact, even certain experiments at relatively small linear plasma devices can be challenging for EIRENE simulations for instance if time-dependent simulations are necessary, whenever the time step imposed by the plasma solver is small compared to the lifetime of the particles as is typical for highly dissipative detached conditions [12]. A good example of that are transient studies of the lithium vapour box experiment at Magnum-PSI aimed at detachment [14].

Generally, plasma ions and electrons are simulated by a fluid approach, and only the neutrals including molecular species are treated kinetically or in hybrid approaches discussed in section 1. In fusion-relevant plasmas, in major parts of the domain the mean free path \( \lambda \) for neutrals is large compared to the gradient lengths \( L \) and the neutral flow is in the large Knudsen number regime, \( K_n = \lambda/L \), for which no accurate fluid closure is available. Hence, a kinetic approach is generally used for neutrals. However, especially in large machines such as ITER and in detached regimes, high-collisional regions (HCRs), where \( K_n \) becomes low, may appear. In these HCRs, the coupling of the neutrals with the background plasma becomes very strong, leading to quasi-Maxwellian distributions for neutrals through e.g. charge exchange reactions with background ions. Typical charge exchange mean-free paths for ITER can be in the mm-range in HCRs in the divertor to m-range in the main chamber, varying over several orders of magnitude mainly due to the inverse dependence on the plasma density. Also neutral–neutral collisions can become
Figure 1. EIRENE-NGM iterative scheme with the CFD codes; the term NGM underlines the role of the code in the coupled packages like SOLPS. (1) The NGM runs simulations for a number of volume cells; (2) the CFD side determines magnetic configuration and runs itself on a grid optimized for it, where the cell shape typically mimics the magnetic field line geometry; this grid may be 3D like in EM3 (a) or 2D, for instance quadrangular plasma cells (b) split into triangles (marked by the star) in SOLPS-ITER; (3) The EIRENE cells, always 3D, correspond to (or approximate) the CFD cells; for 2D CFD cells an extra dimension is provided (magenta stars mark corresponding triangles in 2D SOLPS and 3D EIRENE cells). EIRENE can run fully kinetic or, alternatively, a fraction of neutrals can be treated as fluid providing higher simulation performance; the corresponding fluid calculations are typically provided by the CFD side. In case of APMC hybridisation (see subsection 1.4) no direct coupling with CFD is necessary.

important in HCRs and further reduce the mean-free paths there [12]. This type of situation is computationally demanding in the frame of the MC method, because of the high number of collisions (elementary processes to be simulated) the particle undergoes before being ionized in the domain or absorbed at the surface. HCRs could be more efficiently addressed by using a hybrid fluid–kinetic approach (see section 1).

EIRENE offers many options for simulating the particle trajectories and for estimating neutral sources in the plasma equations. A thorough analysis of the statistical errors and computational cost of these methods has been conducted [15]. To further reduce computational cost, one can turn to advanced fluid neutral (AFN) models, with boundary conditions at the material surfaces derived from the underlying kinetic model. It was demonstrated [16] that these fluid models can be very accurate in HCRs, and at least an order of magnitude faster than fully kinetic simulations. If/when renouncing to the fully kinetic description entirely entails too large modelling errors, a hybrid kinetic/fluid model can be used. Such methods were originally developed in the context of radiation transport [17], and later on also for neutron transport [18] and the Boltzmann–BGK equation [19]. Three approaches were developed for neutral transport in the plasma edge: a spatially hybrid technique (SpH) [20, 21], a micro–macro hybrid method (mMH) [22], and an asymptotic-preserving MC (APMC) scheme [23] (in the latter case no additional fluid model has to be solved). Note that in addition, some ion species may also have to be treated kinetically (see section 2), in particular the often short lived molecular ones. This additional hybridisation of the plasma description is also handled by the kinetic NGM, since molecular ions are important players in the plasma edge chemistry, along atoms and molecules. Lowly charged impurity ions usually do not fully thermalize before being ionized, thus also can be preferred to be treated kinetically. The latter are often simulated by different codes e.g. ERO2.0 [24] in a test-particle approximation using the CFD-EIRENE plasma background as an input. It should be noted that kinetic ions are impacted by the electromagnetic fields and thermal force, thus more data exchange between the NGM trace ion kinetic solver and the fluid side of the CFD-EIRENE package is typically necessary.

The paper gives an overview of the fluid and hybrid fluid–kinetic methods mentioned and suggests the most perspective combinations to be developed. It focuses on the joint effort of EIRENE code developers within the EUROfusion framework aimed at transforming it into a neutral gas module (NGM) suitable for the integrated tokamak modelling. This effort also includes the optimization of the code parallelization by providing an OpenMP–MPI hybrid scheme to the already available MPI approach and some general code refactoring.

1. Fluid–kinetic hybridization

Various methods for hybrid tracking of atomic neutrals (in future applicable also for molecules, ions and molecular ions) in both fluid and kinetic parts of the CFD-EIRENE packages are developed in parallel. To discuss the similarities and differences between the various fluid and hybrid approaches, we consider the following general form of the kinetic equation for the atoms:

\[
\frac{\partial f}{\partial t} + v \cdot \nabla f = S_e + S_t (f),
\]

(1)

where \( f(x, v, t) \) is the possibly 7D distribution function of the atoms. The source \( S_e(x, v, t) = \sum_i \rho_i(x, v, t) \) contains source terms independent of the neutral atom distribution, such as recycling of ions at the different boundaries, volume
recombination, gas puff and also sources of atoms stemming, e.g., from dissociation of molecules are included in this term. The second term on the right-hand side, $S_1(x, v, t) (f(x, v, t))$, describes the impact of collision/absorption events such as charge-exchange (CX) scattering, ionization, and atom reflection at boundaries, which all depend linearly on the distribution function $f(x, v, t)$. For brevity of notation, we will omit the arguments $x, v, t$ in what follows. Using the linearity of the equation, we can write $f = \sum_i f_i$, where $f_i$ is the solution for source $S_i$. Nonlinear neutral–neutral collisions following a BGK [19] approach can be included iteratively in this framework. This is subject of ongoing research, however outside the focus of the present paper. The basic fluid–kinetic hybridization (FKH) methodology will remain the same when including neutral–neutral collisions in a BGK framework, but approaches relying on the linearity of the problem need to be generalized. It should also be noted that neutral–neutral collisions provide an extra relaxation mechanism towards local equilibrium for neutrals, and would thus tend to reduce the Knudsen number (that is, make the neutrals ‘more fluid’). However, while these collisions are important for the overall model accuracy, they are not necessarily dominant in the same spatial regions where CX-collisions lead to the computational bottleneck. This is subject of ongoing research.

1.1. Advanced fluid neutral models (AFN)

A fluid approach is obtained by taking moments of the kinetic equation:

$$\int v \mu(v) \left( \frac{\partial f}{\partial t} + v \cdot \nabla f \right) dv = \int v \mu(v) (S_r + S_1(f)) dv,$$

where $\mu(v) = 1$ for the continuity equation, $\mu(v) = mv$ for the momentum equation (the velocity component parallel to the magnetic field is often the most important), and $\mu(v) = m|v|^2/2$ for the energy equation, with $m$ the particle mass and $\int f_1 \ldots dv$ the integral over the whole velocity space. When assuming a fluid distribution based on a Chapman–Enskog expansion [25] for $f$, we obtain a closed set of equations. For plasma edge neutrals, it is typically a valid approach to reduce the 3D momentum equation to a single momentum equation parallel to the magnetic field, to capture the dominant ion–neutral friction along that direction. For the perpendicular direction and in the absence of plasma drifts, pressure diffusion approximations suffice [26], obtained by neglecting perpendicular inertia terms in the neutral momentum equation. These simplifications are based on a scale-analysis of the terms assuming perpendicular velocities are small compared to the parallel and thermal velocities [26]. Extensions to the context of drifts that become of the order of these velocities are subject of ongoing research. In this project, we have significantly improved the existing fluid models for neutral atoms by deriving the transport coefficients and boundary conditions from the underlying kinetic description (AMJUEL/ HYDHEL databases for the collisional processes and the TRIM database for the reflection physics, assuming an ideal flat surface) [3, 16, 27]. Additionally, an approximate treatment of the effect of molecules is possible through modified boundary conditions. In this way, the transport model and boundary conditions for the fluid neutrals are made consistent with those of the full kinetic simulation, using the same tabulated data, and without introducing ad hoc tuning parameters. In the remainder of the paper, we call these improved models AFNs. Usually, one assumes equal ion and neutral temperatures for a fluid treatment, but reference [28] shows that solving a separate neutral energy equation further reduces the fluid–kinetic discrepancies. With the AFN models, excellent qualitative and quantitative agreement between fluid and kinetic simulations has been achieved for HCRs, without the need for ad hoc parameter tuning in our improved fluid neutral models, both for standalone neutral and coupled neutral–plasma simulations. Moreover, the accurate numerical discretization of the (isotropic) neutral transport fluxes on grids misaligned with respect to the magnetic field was proven essential in this picture [29]. Through the use of boundary conditions based on TRIM data, the impact of wall materials on the reflection properties is now captured by the fluid models [16, 30] — an effect that was inaccessible with earlier fluid neutral approaches.

A pure fluid treatment of the neutral atoms introduces a modelling error that may be quite significant, especially in lower recycling regimes outside of the strict validity range of the models. However, there are obvious benefits in terms of computational cost, with fluid neutral simulations typically being at least an order of magnitude faster than kinetic simulations. The reduced accuracy may be acceptable for initial scoping studies or large parameter scans. If the error is not acceptable, the hybrid methods discussed below can provide an answer.

1.2. Spatial hybridization (SpH)

The validity of the fluid model is restricted to regions of high (mainly CX) collisionality, where the $K_n$ is sufficiently small. This typically occurs only in a small part of a dense divertor. In other regions, in particular the main chamber, the private flux region (PFR), and other remote areas such as sub-divertor structures, kinetic effects need to be incorporated. Various flavours of SpH approaches have been proposed in the literature, usually distinguishing between kinetic and fluid domains, with proper boundary conditions at the interfaces. The exact treatment of the boundary or interface conditions depends on the specific flavour of the hybrid method. Intuitively, the interface has to be placed in a region that is ‘fluid enough’ so that boundary conditions can be defined based on fluid-like distributions, but not ‘too far’ into the fluid so that the computational gain is large. For some methods the interface has to be specified by the user, for others the method does it automatically based on criteria of, e.g., local collisionality to determine evaporation/condensation reactions. In the context of the linear Boltzmann equation, this domain decomposition can be approached in a different way, relying on the localisation of physical sources. Indeed, in the SpH presented in [20], we exploit the linearity of the kinetic equation to split the source term into separate contributions, deciding for each whether to treat the neutrals as kinetic or fluid: $\sum_i S_{i} = \sum_i S_{i, f} + \sum_i S_{i, k}$.
\[ \sum \nu c_{k,i} + \sum \nu f_{i}^{k} \text{ to a corresponding decomposition of the solution } \sum f_{i} \approx \sum f_{i}^{k} + f_{i}^{l}, \]
where the superscript now denotes the treatment, and all sources treated through a fluid model are included in the equation for a single fluid population \( f_{i}^{l} \):

\[ \frac{\partial f_{i}^{l}}{\partial t} + v \cdot \nabla f_{i}^{l} = S_{c,i}^{l} + S_{r,i}^{l} (f_{i}^{l}), \quad (3) \]
\[ \frac{\partial f_{i}^{l}}{\partial t} + v \cdot \nabla f_{i}^{l} = \sum_{i} S_{c,i}^{l} + S_{r,i} (f_{i}^{l}). \quad (4) \]

There are no mutual interactions between the fluid and kinetic parts of the solution. Note that this approach would have to further developed to account for neutral–neutral collisions, which makes the problem nonlinear. Including these collisions can then be achieved by resolving the additional nonlinearity introduced by the BGK model across iterations, as is already the standard procedure for BGK collisions in EIRENE at the moment, but will not be discussed further here. It is immediately obvious that for the source terms \( S_{c,i}^{l} \) and \( S_{r,i}^{l} \) treated kinetically, no modelling error is introduced on the corresponding part of the distribution \( f_{i}^{l} \). This typically concerns source terms due to recycling at main chamber and PFR boundaries, as well as gas puff sources. The sources due to recycling of ions at the targets, as well as those arising from volumetric recombination, are grouped into equation (4), which is then approximated with the AFN models described above, equation (2), typically solved as part of the plasma fluid solver (meaning additional input to EIRENE from the CFD side of the package). At plasma–vacuum boundaries, fluid neutrals can continue their trajectory as kinetic through appropriate interface conditions [20]. The approximation of equation (4) with a fluid model introduces a modelling error on this part of the population. However, in the SpH approach the fluid population is only ‘fed’ with sources in HCRs, where the accuracy of the fluid models is high. The resulting fluid neutral density far from the HCRs is negligible. As a result, the modelling error in the SpH approach is limited, while a significant speed-up is achieved. At present the decision to treat a certain source as fluid or kinetic is still up to the user; methods to automate the decision based on local \( K_{a} \) criteria are being developed.

The SpH method for atoms can easily be coupled to a standard kinetic model for molecules. Thermally emitted particles at a surface are typically launched as \((D_{2})\) molecules, while fast reflected particles are treated as \((D)\) atoms [1]. For fluid atoms impinging on a surface, an estimate of the thermally reflected fraction follows directly from the boundary conditions of the AFN. Upon dissociation, the molecules give rise to another source contribution \( S_{c,i} \), for which it has to be decided whether to treat them (i) kinetically (entering equation (3)); or (ii) as a fluid by adding them as a source term in the fluid neutral model (through equation (4))). These options were explored in [31]. Method (i) is more accurate, whereas method (ii) leads to the largest speed-up. For JET L-mode cases, simulations with the SpH neutral model are on average between 7 and 14 times faster than simulations with fully kinetic neutrals. In that study, the number of MC particles used in the kinetic and hybrid methods was not the same, but was carefully chosen to have the same statistical error on the relevant output quantities of interest, thereby allowing an honest assessment of the speed-up. The hybrid method needs much fewer particles to achieve a given statistical error, since neutrals originating from some of the sources are treated as a fluid (which does not itself introduce additional statistical noise).

In the approach discussed so far, the whole domain is either fluid or kinetic for a given source. In the SpH implemented in [21], with main idea following [32], for all atom sources the distribution function is split into a kinetic and fluid part in the whole domain, so that for a given source fluid and kinetic neutrals co-exist. Additional reaction channels are introduced to represent evaporation/condensation between the kinetic and fluid phases, \( S_{f \rightarrow k} \) and \( S_{k \rightarrow f} \), respectively:

\[ \frac{\partial f_{i}^{l}}{\partial t} + v \cdot \nabla f_{i}^{l} = S_{c,i}^{l} + S_{r,i} (f_{i}^{l}) + S_{f \rightarrow k} - S_{k \rightarrow f}, \quad (5) \]
\[ \frac{\partial f_{i}^{k}}{\partial t} + v \cdot \nabla f_{i}^{k} = S_{c,i}^{k} - S_{r,i} (f_{i}^{k}) - S_{f \rightarrow k} + S_{k \rightarrow f}. \quad (6) \]

By construction, the sum of equations (5) and (6) gives the original kinetic equation (1). Moreover, equation (6) is then approximated through the solution of a fluid model, introducing a modelling error that may also affect the kinetic part of the solution through the coupling terms. At birth, all neutrals are treated kinetically (source \( S_{c,i} \) in equation (5)). The rate coefficients for evaporation/condensation depend on \( K_{a} \) and are chosen such that the trajectories of kinetic atoms are quickly stopped when entering a fluid region and contribute to the source of fluid atoms there. For the detailed transition criteria, we refer to reference [21]. The expected computational gain comes from this shortening of the trajectories in HCRs. The implementation of these models is straightforward and does not require modifications of the kinetic solver. However, the transition between fluid and kinetic regions/boundaries is based on ad hoc criteria, whose choice affects the accuracy of the scheme. If the \( K_{a} \) dependence of the condensation/evaporation rates is sharp enough so that the transition between domains where the fluid (resp. kinetic) phase strongly dominates occurs over a few grid cells, this method can be seen as an immersed boundary implementation of a standard SpH technique. Such a sharp transition ensures the ‘quick’ termination of kinetic particle trajectories when entering into fluid regions thereby optimizing the speed-up.

As future work, we plan to combine both SpH methods, by investigating optimal combinations of source splitting and evaporation/condensation reactions. For both SpH methods, the kinetic parts of the distribution \( f_{i}^{k} \) are positive by definition and represent a real part of the distribution for the total neutral population. The corresponding neutrals are simulated with a regular MC procedure, requiring limited implementation effort. The disadvantage of the SpH methods is that they are not completely equivalent with the kinetic description, but introduce a model error. The first SpH method is only accurate if the fluid limit is reached near the sources treated as fluid. The accuracy of the second SpH method strongly depends on the spatial distribution of the condensation and evaporation.
sources. These disadvantages disappear when using a mMH approach as introduced in the next section.

1.3. Micro–macro hybrid (mMH) approach

In the mMH approach of reference [22], the (complete) neutral distribution function $f$ is split into a fluid part $f^f$ and a ‘kinetic correction’ $f^δ$ seamlessly in the entire simulation domain, such that $f = f^f + f^δ$ exactly holds. A consistent set of equations is derived in such a way that the sum of the fluid part and the kinetic correction gives exactly the same solution as the original fully kinetic equation. For the fluid part of the distribution, this leads to the AFN equations, with additional corrections on source and transport terms due to the kinetic correction population. The kinetic correction is obtained from a modified kinetic simulation involving positive and negative correction particles, with net zero density and momentum and provides an exact closure for the AFN model. As a result, the method requires more substantial development efforts for the kinetic correction terms, and requires a fluid solver acting for the whole volume right down to the first wall [33]. The benefit of mMH is that the solution is equivalent to the solution of the fully kinetic equation and independent of the recycling regime, illustrated in figure 2 for a JET-like slab case. However, for a significant computational benefit, the distribution implied by the fluid model should already be close to the fully kinetic distribution (hence, the kinetic correction part should be small), which is expected only for high-recycling or detached conditions. In regions of the simulation domain where kinetic effects are important, substantial cancellation errors arise, that presently hamper the efficiency and convergence of the method [34]. Note that the efficiency (speed-up) for individual source estimation on fixed plasma background is typically highest for the particle source, and lower for the higher order velocity moments (momentum/energy sources) which are more strongly impacted by kinetic effects [34]. When looking at the coupled plasma–neutral simulations, this will lead to a single ‘effective’ speed-up over the entire simulation. To alleviate these effects, further research will look into projection techniques to limit the cancellation errors [22], as well as a synergistic combination with the SpH approach where the mMH technique is not applied to the complete neutral distribution, but only to the part of the distribution treated as fluid in the SpH approach. In this case, $f = \sum_i f_i = \sum_i f^k_i + f^f + f^δ$ is again an exact decomposition. The terms $f^k_i$ that show predominantly kinetic behaviour are treated with a standard kinetic simulation, whereas the mMH approach will provide an exact closure for the part of the neutral population treated as fluid.

Although both are MC simulations, the nature of a kinetic correction simulation needed by the mMH approach is substantially different from an MC simulation of the ‘regular’ kinetic problem of equation (1). Because the largest share of the CX collisions is captured by the fluid model, absorption events are much more dominant in the kinetic correction equation, while scattering (CX) events dominate the regular kinetic problem. A rigorous analysis of optimal combinations of MC tracking and source estimation schemes demonstrates that non-analog track-length or next-event tracking schemes are optimal under such conditions [15] while the default treatment in EIRENE is an analog particle tracking procedure. Further speed-up of the mMH approach can be expected by using these optimal tracking procedures [34]. The insights from [15] may also serve to optimize regular kinetic simulations by automatically selecting the best combination of tracking and estimator schemes based on local collisionality.

1.4. Kinetic-diffusion Monte Carlo method (KDMC)

The mMH and SpH approaches both require an additional discretization of the spatial domain suitable for the fluid
An alternative that avoids these complications is the *kinetic-diffusion MC scheme (KDMC)*, an asymptotic-preserving scheme which is fully MC (APMC) [23]. In this method, the MC particles themselves behave alternately according to the kinetic equation and according to the fluid limit. During kinetic phases, the MC particles are simulated with a standard particle tracing method. During the fluid phases, the MC particles move via a random walk method, an MC discretization of the limiting fluid equation. In the KDMC-scheme, a random walk step is used after each collision during the kinetic phase. This random walk step continues until the end of the time step, the size of which is the numerical parameter of the scheme. The coefficients of the random walk step are designed to such that the positional increments match the kinetic process’s mean and variance exactly in every time step. Furthermore, the correlation of the motion between subsequent time steps is maintained by the proposed combination of kinetic and fluid-like parts. Outside of the truly kinetic or truly fluid regimes, the KDMC scheme results in a bias. This small bias in the simulation determined by the time step is removed with a newly developed multilevel extension of the algorithm, where the multiple time step sizes define the levels [15, 35]. The final challenge for this method is the design of estimators for the random walk steps that capitalize on the achieved speed-up. Such estimators have been developed for a 1D simulation, and are being extended to higher dimensions.

### 2. Kinetic ions

In the previous section, several approaches designed to treat neutrals at least partially as a fluid in order to reduce the computational cost have been presented. Here we discuss the opposite situation, in which ion species are treated partially in a kinetic approach within the EIRENE solver. This is relevant for e.g. molecular ions involved in plasma edge chemistry, but also potentially for lowly ionized impurity ions. In fact, the relaxation time for the velocity distribution of the latter towards a local Maxwellian (through Coulomb collisions) can be longer than their ionization time. In such cases, where features of the neutral impurity velocity distribution can propagate to ions, only a kinetic approach can capture these effects, which should impact the parallel transport force balance for impurities. Therefore, a kinetic approach for plasma ions at trace levels is developed inside the EIRENE code. In fact, treating ions on the kinetic side of CFD-EIRENE packages in addition to a fluid/hybrid approximation for neutrals provides more flexible and seamless coupling between the codes as well as an internal benchmarking mechanism. In the last two years fundamental physical enhancements of the kinetic ion transport part of EIRENE were performed [36], by adding first-order drift effects, cross-field diffusion, and magnetic mirror force. These additions, which are relevant for thoroughly investigating the full three-dimensional influence of impurities on actual fusion devices, have been cross-checked on simple model cases against analytical properties of passing and trapped (banana) particle orbits, as well as checking on the introduction of numerical diffusion by our integration scheme. These features are still under development in preparation for use in full physics cases.

### 3. Parallelisation of EIRENE

The MC method implemented in EIRENE follows independent particles (in between iterations with a CFD code), even in non-linear cases involving collisions between tracked species, where the code relies on an iterative BGK approach [19]. Massive parallelisation is thus in principle straightforward, as it involves communications only before and after the MC calculation itself. For a given accuracy (statistical noise level) parallelisation will reduce the wall clock time, or for a given wall clock time the accuracy can be improved. Reducing the statistical noise by one order of magnitude requires 100 times more test particles because of the square-root scaling characteristics of MC approaches. Because of the stratified sampling used in EIRENE (where one stratum often corresponds to a physical source, see section 1), different strategies are possible: the strata can be treated sequentially with particles distributed on all cores for each strata; or the strata can be treated in parallel, each one being allocated a number of cores on the basis of predefined criteria (involving the number of particles and the relative flux associated to a strata). The latter reduces the amount of communications but load balancing must be carefully considered. Both strategies were available in EIRENE and rely on a distributed memory parallelization using the MPI library. However, for large 3D grids used for transport calculations (e.g. EMC3-EIRENE) on large machines or for turbulent calculations on present day devices the memory requirements to store the ~100 input and output tallies of EIRENE become large enough so as to preclude running with one MPI process per core on a node. Moreover, when coupled to hybrid OpenMP/MPI parallelized codes such as SolEdge3X [11] the current implementation leads to an inefficient CPU resource usage when running coupled to EIRENE and running with more than one thread per core. As a result, an OpenMP shared memory layer has been added into EIRENE, where all the particles attributed to an MPI process are distributed among threads on the local node. This allows efficient use with hybrid OpenMP/MPI codes and has the potential to reduce the memory consumption by a factor of at most $1/N_{\text{core}}$, where $N_{\text{core}}$ is the number of cores on the node. All tallies are treated as shared arrays within threads for a given MPI process and are therefore not duplicated in memory. In practice, the actual reduction in memory usage comes close to this limit as illustrated in figure 3(b), where using 48 cores implies only ~10% increase in memory while the same run with 48 MPI process implies a ~48 fold increase. However, figure 3(a) shows that the parallel efficiency strongly decreases above 12 cores on this architecture (Marconi-fusion [37]), because of the need of ensuring correctness when updating shared tallies (using ATOMIC clauses), but also because of memory access bottlenecks. Indeed, the current structuring of tallies in memory allows strong flexibility but is not optimal w.r.t. to OpenMP (or as well for usage of GPU-based systems which are typically also with shared memory), and addressing these shortcomings...
Figure 3. OpenMP/mpi hybrid parallelisation of EIRENE-NGM demonstrated on the ITER test case. The OpenMP efficiency (left, (a)) diminishes with the number of cores/threads, however the drop is monotonic and consistent. On a positive side OpenMP provides dramatic reduce of the peak memory consumption (right, (b)), which is critical for large and detailed simulation cases in particular in a view of the upcoming ITER and DEMO predictive modelling.

would require a major code restructuring. On the longer term, further relaxing memory constraints will require implementing a domain decomposition approach, with nodes being allocated chunks of the full grid, and particles distributed among threads on the node.

4. Discussion and perspectives of FKH approaches

A summary of the main advantages and remaining issues of the various fluid and hybrid methods is presented in table 1. Figure 4, gives an overview for the hybridisation methods discussed in this paper, in terms of the balance between computational speed (increases towards the left of the figure) versus modelling accuracy (increases towards the right), which are the main performance criteria for hybrid kinetic–fluid approaches, and typically need to be balanced against each other. To find this balance, one needs to investigate how both the accuracy and computational cost of each of the methods scales with the method parameters. Then, one needs to decide on either computational budget and choose those method parameters that maximise accuracy, or, conversely, decide on a desired accuracy and choose those method parameters that minimise the computational cost. It is quite natural to expect the conclusion to depend on both the case studied and the desired accuracy (or the available computational budget). Hence, a detailed comparison of different hybrid schemes that leads to meaningful conclusions is a difficult challenge. Here, we restrict ourselves to indicating the factors that influence the comparison, and overview a systematic way of optimising method parameters for each of the hybrid methods.

Computational errors in the fluid–kinetic context generally consist of discretization errors, cancelation errors, statistical errors, biases, and modelling errors. Except for modelling errors, these errors can be lowered by adjusting the method parameters, hereby increasing the computational cost. For instance, for the discretization errors, this is done by refining the mesh. For the statistical errors and biases, increasing the number of samples reduces the error [38].

AFN, SpH, and mMH cope with discretization errors for the respective fluid parts. To optimally choose grid sizes, guidelines are discussed in [38], on a coupled plasma–neutral level. The guidelines provided there also apply for the neutral simulation on its own: by using a sequence of coarsened/refined grids, the size of the discretization error can be estimated via Richardson extrapolation. Additional errors present in the particle (MC) part of SpH and in mMH and KDMC that have to be taken into account are the errors arising due to stochasticity. In a simple coupled plasma–neutral simulation, there is on the one hand a statistical error that scales as $1/\sqrt{N}$ with $N$ the number of particles. On the other hand, a bias due to the nonlinear coupling with the plasma simulation is present that scales as $1/N$. Simulation strategies with averaging over different runs, results in different scalings, as shown and discussed at length in [39].

In [38] a methodology is presented that jointly selects the grid size and the number of particles optimally for a given computational time. The modelling error present in AFN and SpH for instance, depends critically on which application case is considered and, in the case of SpH, also on how the boundary between the submethods is defined. The truly hybrid methods such as SpH, mMH, and KDMC benefit from a large scale separation between different regions of the application case, meaning that there are some clearly kinetic regions and some clearly fluid regions.

All methods have difficulty with regions of intermediate, often called rarefied, behaviour. The precise balance struck between scale separation, accuracy, and code performance, is a key point of further investigation. In this text, we will only go over the key effects. For SpH, this intermediate region has to be located and the boundary between its two submethods has to be established there. If the boundary is taken close to the truly fluid region, the resulting model error will be low, but at the cost of a sharp increase in computational cost, and vice versa if the boundary is close to the truly kinetic region. The mMH method will automatically result in a fully accurate simulation of the intermediate region, but if the scale separation is low, this might result in very poor code performance in
Table 1. An overview of implemented FKH approaches.

<table>
<thead>
<tr>
<th>FKH approach</th>
<th>Main idea and parameter(s)</th>
<th>Advantages</th>
<th>Issues</th>
<th>Development status and performance gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced fluid neutral models</td>
<td>Replace kinetic simulation with a fluid model, tailored to conditions of high CX collisionality</td>
<td>- Large speed-up compared to kinetic simulation (&gt; order of magnitude)</td>
<td>- Introduces a modelling error, that may be substantial in conditions/regions</td>
<td>- Implemented in SOLPS-ITER - Good accuracy for high recycling conditions, incl. ITER, demonstrated</td>
</tr>
<tr>
<td>(AFN) [16]</td>
<td></td>
<td>- No statistical noise</td>
<td></td>
<td>[27, 30]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Tight coupling with plasma equations may improve convergence</td>
<td></td>
<td>- Speed-up factor &gt; 10</td>
</tr>
<tr>
<td>Spatial (SpH)-based on source</td>
<td>Treat neutrals born in high $K_n$ regions as kinetic, and born in low $K_n$ regions as fluid</td>
<td>- Straightforward implementation and coupling to molecules</td>
<td>- (Small) remaining modelling error - Choice of kinetic/fluid source treatment up to user</td>
<td>- Implemented in SOLPS-ITER - Good accuracy for high recycling conditions, incl. ITER, demonstrated</td>
</tr>
<tr>
<td>location [31]</td>
<td></td>
<td>- Clear improvement in accuracy compared to AFN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpH with evaporation /</td>
<td>Co-existence and interaction of two phases (kinetic and fluid) on the full domain, can be made</td>
<td>More seamless transition, automatic procedure for spatial</td>
<td>- Sub-optimal speedup because kinetic trajectories entering HCR not terminated</td>
<td></td>
</tr>
<tr>
<td>condensation [21]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>micro-Macro (mMH)</td>
<td>Based on exact decomposition of kinetic equation in fluid and kinetic correction parts</td>
<td>Modelling error can be completely removed, at all collisionalities</td>
<td>Substantial - Implemented in low collisional /kinetic regions hamper convergence - Requires complete overlap of grids for fluid and kinetic correction neutrals grid (up to the wall)</td>
<td>- Implemented in SOLPS-ITER - Speed-up factor 5-10 demonstrated in simplified geometries</td>
</tr>
<tr>
<td>APMC option:</td>
<td>Particles follow a hybridized path that combines advection diffusion and kinetic steps</td>
<td>- No need to resolve individual collisions - No need for separate fluid/neutral grid - free of cancellation - bias can be suppressed by ML-KDMC</td>
<td>Currently only available for single species scattering/absorption</td>
<td>The approach is for now just demonstrated on simplified problems, treatment of simulation cases similar to e.g. SOLPS-ITER tasks is in elaboration.</td>
</tr>
<tr>
<td>Kinetic diffusion (KDMC) [23]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>multilevel approach (ML-KDMC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| KDMC method, a solution to this issue has been developed by incorporating the KDMC method into the multilevel MC framework [35], which combines simulations with different time steps. The cheap, large time-step simulations allow to quickly reduce the statistical error. The expensive, small
Table 2. Alternative ways to increase performance.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Main idea and parameter(s)</th>
<th>Combinability with FKH?</th>
<th>Development status and performance gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parallelisation (brute force approach)</td>
<td>- Strong scaling allows reducing the wall clock time or improving the MC statistics - Weak scaling allows more detail and larger simulation volume (neutrals)</td>
<td>Provide additional performance on the remaining kinetic calculation (largest effect expected on APMC where no fluid solver is required for)</td>
<td>- Good MPI scaling - Decent OpenMP efficiency within 10-20 threads/cores - About 100 times reduction of the memory peak consumption.</td>
</tr>
<tr>
<td>Optimised tracking of kinetic processes</td>
<td>Selection of simulation/estimator procedure based on analysis of variance</td>
<td>- Combination with mMH and SpH straightforward</td>
<td>Results in prototype code, can be transferred to EIRENE</td>
</tr>
<tr>
<td></td>
<td>Improve CRM performance by reorganisation and scalings for e.g. isotopologically resolved hydrogen data</td>
<td>The same reaction rates can be utilized by both the fluid and kinetic parts of FKH codes</td>
<td>- Schemes for CRM construction exist. - Isotopologically resolved hydrogen data is still scarce</td>
</tr>
</tbody>
</table>

Figure 4. Hierarchy of neutral modules (above the dashed line) and hybrid fluid–kinetic approaches already implemented in EIRENE (below the dashed line).

time-step simulations allow to recover the full kinetic accuracy.

Besides the main criteria of accuracy and code performance, there are several more qualitative aspects that differentiate the methods. Methods consisting of multiple methods such as SpH and mMH are more difficult to implement than KDMC and AFN, but also the incorporation of those methods into EIRENE may require significant changes to the existing code. The purely particle-based method of KDMC, is furthermore trivially parallelizable. Other methods have the advantage of being more mature in their development such as SpH and AFN. In conclusion, the decision on which of the many existing methods is the best choice in the long term is still open. In this section, we stress the need for sound choices in grid size and particle numbers as in [38] and we identify the error-accuracy balance in rarefied regimes for SpH, mMH, and KDMC as an important research topic.

As discussed above, further research will investigate synergistic combinations of the various hybrid approaches, aiming for at least an order of magnitude speed-up for the most accurate hybrid schemes. To our view, we need to keep two FKH tracks under development: one based on AFN/SpH/mMH and another based on the asymptotic-preserving KDMC scheme. The first track should evolve to a single unified option combining the two SpH options via the use of the condensation/evaporation terms, and providing exact closure for the fluid neutral population through an mMH approach. This line is already strongly developed, however further development
is needed specifically to limit the cancellation error in the mMH scheme. KDMC accuracy is compromised in a similar way by a bias error at intermediate collisionalities, which however can be completely suppressed by the multilevel ML-KDMC. In the end, the final performance and flexibility on DEMO scale will be decisive. In addition, the first (AFN) track will always enable tighter coupling with the plasma equations (could be beneficial for convergence of the coupled system) and should have lower MC noise (beneficial for sensitivity analysis as used e.g. in gradient-based optimization studies [40, 41]). The KDMC approach, on the other hand, allows standalone use of EIRENE without iterations with the CFD code, which is beneficial for geometry flexibility and following of A & M processes including kinetic ions.

Table 2 shows alternative ways to increase performance, which should work independent on the simulation type—hybrid or full kinetic except to optimisations. One can expect also some performance gain from general restructuring and refactoring of the code, including the use of the modern, HPC-friendly output format HDF5, providing e.g. parallel writing by multiple cores. Also the accurate computation of sensitivities in the presence of MC noise (adjoint approach) has been demonstrated [43], and will be pursued in further EIRENE development involving algorithmic differentiation as shown in [41] for the (deterministic) plasma solver within SOLPS-ITER.

5. Summary and conclusion

FKH approaches are developed (for instance SpH and mMH gaining performance due to a partial use of the AFN approximation [16]) for the CFD-EIRENE packages [20, 21, 34]. They combine improved computing performance with model accuracy approaching full kinetic simulations. The alternative perspective APMC approach is also considered [23], including development and first tests of the newly proposed KDMC formulation with a multilevel option (ML) [35] aimed to overcome the bias. In addition, the option to track ions kinetically is improved [36]. The advantages of hybridisation methods are compared based on experience from the first applications to test cases relevant for fusion devices. Currently, the main effort is on (1) basic development of the approaches (2) validation with full-kinetic simulations to determine the gain in computational speedup and optimal parameters (3) impact demonstration of new physics included on, for example, ITER-relevant applications (4) unification of the A&M processes including kinetic ions.

At the present stage, we recommend to continue development of 2 FKH lines in parallel: the merged into one AFN/SpH/mMH approach (the way for that is suggested) and the ML-KDMC approach [23]. Each of the approaches is expected to have an automatic accuracy control and self-adaptation mechanism for any case at hand.

The hybrid OpenMP–MPI code parallelization and optimisation of the A & M process treatment (improved CRMs) go mostly in parallel adding an additional factor to the improvement of the EIRENE-NGM performance. However, this factor can depend on the final selection of the FKH scheme and overall optimisation of the code.

Acknowledgments

The paper presents a joint effort within the EUROfusion Theory and Advanced Simulation Coordination (E-TASC), task ‘NGM’ 2019–2020 and the related O-EIRENE HLST project. This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014–2018 and 2019–2020 under Grant Agreement No. 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission. W. Van Uytven is funded by a PhD fellowship of the Research Foundation Flanders (FWO). Parts of the work were supported by the Research Foundation Flanders (FWO) under project Grant G078316N.

ORCID iDs

D.V. Borodin https://orcid.org/0000-0001-8354-1387
F. Schluck https://orcid.org/0000-0003-4634-0679
S. Wiesen https://orcid.org/0000-0002-3696-5475
S. Brezinsek https://orcid.org/0000-0002-7213-3326
W. Dekeyser https://orcid.org/0000-0003-1954-3729
S. Carli https://orcid.org/0000-0001-6581-2745
M. Blommaert https://orcid.org/0000-0002-2616-8042
W. Van Uytven https://orcid.org/0000-0002-7253-2087
M. Baelmans https://orcid.org/0000-0001-9905-4583
B. Mortier https://orcid.org/0000-0003-1788-7819
G. Samacé https://orcid.org/0000-0001-8433-4523
Y. Maranet https://orcid.org/0000-0002-9970-538X
E. Westerhof https://orcid.org/0000-0002-0749-9399
J. Gonzalez https://orcid.org/0000-0001-7905-5001
M. Groth https://orcid.org/0000-0001-7397-1586
A. Holm https://orcid.org/0000-0002-8391-8364

References


[9] Bufferand H. et al. 2015 Numerical modelling for divertor design of the WEST device with a focus on plasma–wall interactions Nucl. Fusion 55 055025


[16] Horsten N., Samaey G. and Baelmans M. 2017 Development and assessment of 2D fluid neutral models that include atomic databases and a microscopic reflection model Nucl. Fusion 57 116043


[34] Horsten N., Dekeyser W., Blommaert M., Samaey G. and Baelmans M. 2020 A hybrid fluid–kinetic neutral model based on a micro-macro decomposition in the SOLPS-ITER plasma edge code suite Contrib. Plasma Phys. 60 e201900132


[38] Ghoos K., Börner P., Dekeyser W., Kukushkin A.S. and Baelmans M. 2018 Grid resolution study for partially detached ITER divertor plasma Nucl. Fusion 59 026001


[40] Baelmans M., Blommaert M., Dekeyser W. and Van Oevelen T. 2017 Achievements and challenges in automated parameter, shape and topology optimization for divertor design Nucl. Fusion 57 036022

