



l'IDRIS

Numéro 2 - Avril 1996

Édito

L'objectif du CNRS lors de la création de l'IDRIS était la mise en place d'un centre de ressources et de compétences diversifié, polyvalent et évolutif, doté d'architectures scalaires, vectorielles et massivement parallèles.

Qu'en est-il aujourd'hui ?

LE CALCUL SCALAIRE

Le calcul scalaire s'effectue sur une grappe de stations IBM RS 6000 très puissantes. Cette plate-forme a fait beaucoup d'adeptes. En effet, l'utilisateur peut pendant 48 heures prendre le contrôle exclusif d'une station puissante, ce qui n'est pas souvent possible avec les matériels des laboratoires. Ces stations possèdent un autre atout non négligeable : les logiciels y sont déjà installés.

LE CALCUL VECTORIEL

Le calcul vectoriel est le noyau dur du calcul intensif de haute performance fait à l'IDRIS et s'effectue sur les plates-formes Cray C98 et C94. La capacité maximale de production de ces machines est de 90.000 heures monoprocesseur par an (h/a). La demande se stabilise autour de 125.000 heures/an et la production pour la recherche autour de 80.000 heures/an. Le calcul vectoriel rassemble 450 projets scientifiques (dont une partie importante a des retombées internationales, européennes ou industrielles) et 1500 utilisateurs. Le rapport de conjoncture préparé par le Conseil Scientifique de l'IDRIS examine en détail les résultats scientifiques obtenus en 1994/95 sur nos supercalculateurs vectoriels. Il souligne l'excellente adaptation de l'environnement vectoriel actuel aux besoins de nombreux sujets de recherche et son impérative

nécessité pendant quelques années encore.

LE CALCUL PARALLÈLE

Le calcul parallèle a connu un démarrage assez tardif puisque nous avons préféré privilégier la mise en place d'un environnement de stockage des données performant. Un T3D (128 nœuds, 8 gigaoctets de mémoire) est en exploitation depuis septembre 1995 et vous trouverez les commentaires sur les premières expériences des utilisateurs dans ce journal. Un T3E (256 nœuds, 32 gigaoctets de mémoire) attendu sur notre site à la fin de l'été prochain, ouvrira les portes d'une nouvelle physique inaccessible par les plates-formes vectorielles. Environ 110 projets sont aujourd'hui en cours et une trentaine sont déjà très actifs. Le T3E accroîtra les ressources parallèles de manière significative et celles-ci seront à terme pleinement utilisées.

L'ENVIRONNEMENT ACTUEL

Aux machines de calcul s'ajoutent les différents serveurs : le serveur de données, en forte évolution cette année à cause de l'arrivée du T3E, et le serveur de visualisation présenté dans ce journal. L'environnement actuel de l'IDRIS semble donc s'adapter très correctement aux besoins d'une recherche d'avant-garde. Notre mission future sera de préserver cette harmonie fine et délicate entre une industrie des supercalculateurs - qui se cherche encore - et les besoins de la recherche scientifique. Il nous faudra garantir à nos utilisateurs le changement mûr et réfléchi avec une évolution permanente et soutenue de la qualité des logiciels (systèmes d'exploitation, compilateurs, aides à la programmation) de nos supercalculateurs.

V. Alessandrini

SOMMAIRE

- EDITO = P1
- T3D = P2/3
- 10 POINTS SUR PVM = P4

- DU CÔTÉ DE LA VISUALISATION = P5
- NEWS = P6



la demande du Comité des Utilisateurs, une journée T3D a été organisée à l'IDRIS le 14 mars dernier. Elle devait permettre de dresser le bilan de l'utilisation de la machine grâce à l'expérience de ses utilisateurs. Six mois après sa mise en service, le T3D (128 processeurs, 8 gigaoctets de mémoire) est quasiment en pleine charge. Certains projets sont déjà en phase de production et d'autres en phase finale de développement.

COMPTÉ-RENDU

Denis Girou (coordinateur du support au parallélisme à l'IDRIS) fait une brève présentation du T3E attendu sur le site l'été prochain et cède ensuite la parole aux utilisateurs.

Plusieurs responsables de projets exposent ensuite tour à tour leurs motivations scientifiques, leurs choix algorithmiques, leurs stratégies de parallélisation et optimisation des codes et enfin les premières conclusions sur le T3D en tant qu'outil de production scientifique.

- Yves Epelboin (laboratoire de Minéralogie-Cristallographie, Paris VI - Jussieu) : présentation de sa stratégie dans la mise au point d'une application parallèle dans le domaine de la diffraction des rayons X. Un code portable écrit en PVM a été aisément adapté et optimisé pour le T3D. Ce type de code pourrait bénéficier d'une taille mémoire par processeur plus importante que celle disponible aujourd'hui. L'arrivée prochaine du T3E assurera le passage en production.

- Marie Madeleine Rohmer et Marc Benard (laboratoire de Chimie Quantique, Strasbourg et ASCI, Orsay) : parallélisation du code de chimie quantique ab initio ASTERIX. Des algorithmes plutôt sophistiqués permettent d'optimiser le calcul d'un grand nombre d'intégrales bioélectroniques. Ce code, destiné à l'analyse des grosses molécules qui échappent à la puissance du C98, montre une excellente efficacité de 2 à 64 processeurs.

- Maurice Meneguzzi (ASCI, Orsay) : expérience du portage de simulation d'écoulements diphasiques dans une conduite avec des bulles ou des particules de taille finie. Écrit en C, ce code fait appel à MPI pour les communications. Une première version parallélisée de CHARMM a été utilisée par Jean Durup (laboratoire de Physique Quantique, Toulouse) pour avancer le calcul de la force d'hydratation entre deux protéines en milieux aqueux. Ce code nécessite l'usage de la bibliothèque SHMEM et bénéficie d'une bonne efficacité jusqu'à 32 processeurs.

- Rémi Jacques, Patrick Le Quéré et Olivier Daube (Limsi, Orsay) : simulations numériques directes d'écoulements turbulents dans les cavités interdisques. Ce code fait appel à des algorithmes multidomains et obtient des performances similaires sur le T3D que l'algorithme monodomaine sur le C98. Les points de discrétisation accessibles sur le T3D sont cependant plus importants que sur le C98 en conservant des temps de calcul raisonnables.

- Anne Héron et Jean Claude Adam (École Polytechnique) : parallélisation et optimisation d'un code particulaire relativiste destiné à l'étude de l'interaction laser-plasma. Elle implique la parallélisation d'un système hyperbolique (les équations de Maxwell), d'une équation elliptique (l'équation de Poisson) et des équations de mouvement comportant un nombre élevé (typiquement 107) de particules par une technique de décomposition des domaines. L'utilisation de la bibliothèque SHMEM assure d'excellentes performances parallèles. Certaines optimisations permettent d'améliorer très nettement une version initiale assez médiocre sur le plan des performances. Après optimisation, le code devient aussi efficace sur 8 processeurs du T3D que sur 1 processeur du C98. Les auteurs concluent sur l'effort d'optimisation, qui devrait assurer un gain de facteur 2 par rapport au code optimisé sur une machine vectorielle.

- Maurice Imbard (LODYC, Paris VI - Jussieu) : version parallèle du modèle OPA de circulation de l'océan. Ce modèle s'appuie sur les équations de Navier-Stokes simplifiées et est utilisé depuis plusieurs années dans la communauté scientifique « Environnement ». Des machines parallèles comme le T3D/T3E permettent aujourd'hui d'accroître la résolution du modèle, étape obligatoire pour représenter la dynamique de l'océan dans le cadre des études réalisées en océanographie et en climatologie. Le modèle, dans sa version océan mondial très utilisée dans les études du climat, a été intégré après plus d'un an de simulations. 40 processeurs sont maintenant nécessaires pour égaler les performances obtenues par un processeur C98 à quantité égales d'entrées/sorties.

- Victor Alessandrini (IDRIS) : parallélisation et optimisation d'un code destiné à l'étude de la chromodynamique quantique sur réseau. Il nécessite la décomposition en domaines d'un maillage régulier à 4 dimensions. Les algorithmes choisis permettent d'effectuer une décomposition totalement générale, définie à l'exécution, sans modification ou recompilation du code. Ceci entraîne des communications à grains fins, optimisées par la librairie SHMEM. Toutefois, le code est parfaitement extensible (« scalable ») : en passant de 8 à 64 processeurs, le nombre de degrés de liberté augmentent d'un facteur 8 et les temps d'exécution restent identiques, à 1% près. L'amélioration de la performance apportée par le générateur de nombres aléatoires de la Benchlib est particulièrement impressionnante. Cette expérience montre que le T3E promet d'être une machine très compétitive pour ce type d'étude.

- Jean-Yves Cognard (laboratoire de mécanique et Technologie, Cachan) : expérimentation d'une technique de sous-structuration à l'aide de l'atelier logiciel des éléments finis CASTEM 2000 développé au CEA Saclay (PVM étant utilisé pour les échanges de messages). Au démarrage du projet, une version prototype sur SP2 et sur réseau de stations utilisant PVM était disponible. L'objet de l'étude portait sur le comportement de l'algorithme pour des simulations à grand nombre de degrés de liberté en comparant avec le SP2. Des difficultés sont apparues dans le portage de ce code et notamment pour les entrées/sorties.

REMARQUES ET CONCLUSIONS

Le T3D est perçue comme une machine robuste et fiable.

Deux points forts : adressage global de la mémoire et très faible latence des communications qui permettent le développement des codes avec un haut niveau de parallélisme.

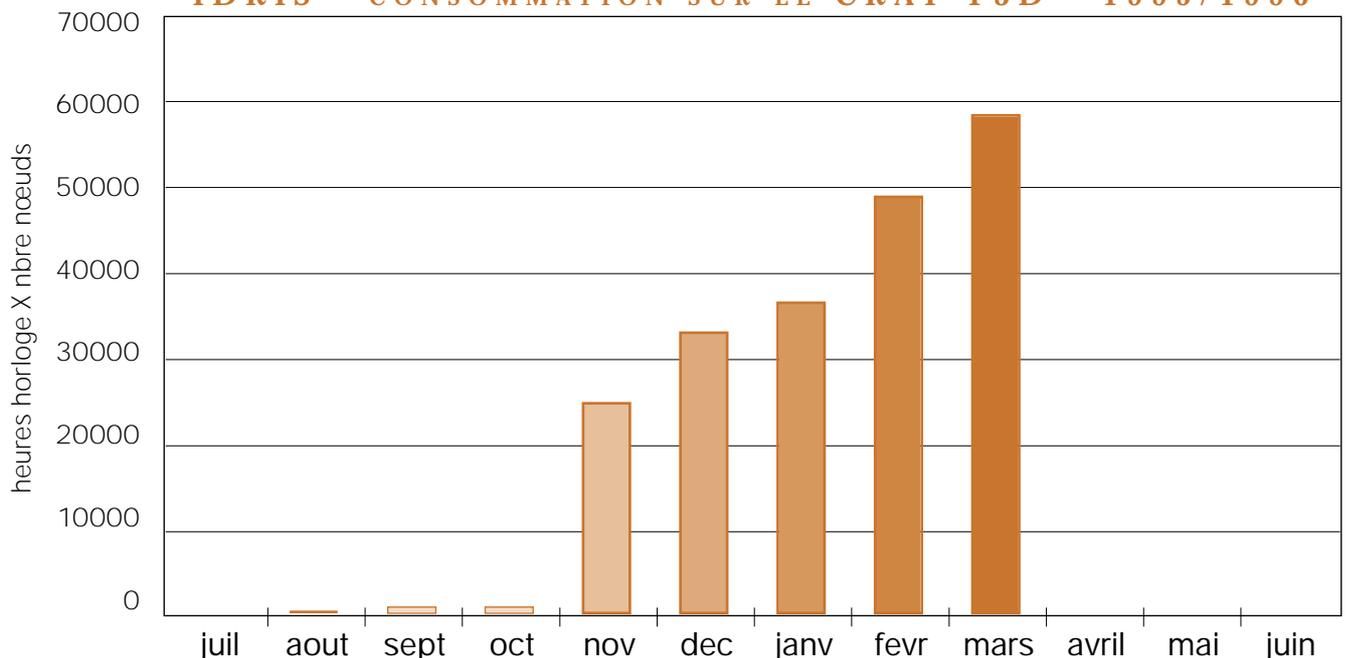
Deux points faibles : la puissance limitée des nœuds de calcul et les performances d'entrées/sorties qui devraient être améliorées de manière significative sur le T3E.

Les utilisateurs engagés dans le portage et l'optimisation des codes ont tenu à insister sur l'efficacité du support au parallélisme apporté par les ingénieurs de l'IDRIS.

Cette journée a permis de mettre en exergue l'implantation du parallélisme au sein de la communauté scientifique. Après 6 mois d'exploitation, le T3D est pleinement utilisé. Une nouvelle culture voit le jour : les machines parallèles deviennent aujourd'hui un outil efficace de production scientifique.

Beaucoup de laboratoires et d'organismes de recherche ont été les précurseurs du développement du parallélisme massif en France. Mais nous constatons que le projet de parallélisme du CNRS porté par l'IDRIS accélère aujourd'hui significativement la diffusion de cette technologie.

IDRIS - CONSOMMATION SUR LE CRAY T3D - 1995/1996



Dix points à connaître sur PVM

1 • Qu'est-ce que PVM ?

PVM (Parallel Virtual Machine) est une bibliothèque de fonctions permettant la mise en œuvre d'applications parallèles sur machines à mémoire distribuée, par échange explicite de messages. La machine parallèle utilisée peut être virtuelle (exemple : grappe de stations hétérogènes).

2 • Comment peut-on s'informer sur PVM ?

L'IDRIS a mis en place un cours concernant la programmation avec PVM sur T3D/T3E. Vous pouvez connaître les dates des cours ou récupérer les supports en consultant le serveur Web de l'IDRIS (<http://www.idris.fr>). De plus, voici quelques livres de référence :

- PVM3 User's Guide and Reference Manual (<http://www.netlib.org/pvm3/ug.ps>)
- PVM : Parallel Virtual Machine, MIT Press (<http://www.netlib.org/pvm3/book/pvm-book.html>)

3 • Quel est le principe de PVM ?

La transmission d'information entre plusieurs tâches s'exécutant en parallèle se fait explicitement par échange de message. Lorsqu'un processus envoie un message (*send*), un autre doit le recevoir (*recv*). La synchronisation des processus est, elle aussi, explicite. Elle est donc à la charge du programmeur.

4 • Quels modèles d'exécution sont compatibles avec PVM ?

PVM supporte différents modèles d'exécution. Le choix est plutôt guidé par nos habitudes de programmation et le modèle supporté par l'architecture de la machine parallèle cible. Sur T3D, les processus exécutent le même programme sur des données différentes (SPMD). Sur T3E, nous aurons accès à la fonction de génération de processus (*sapwn*) comme sur grappe de stations. Les processus pourront donc exécuter en parallèle des programmes différents (MPMD).

5 • Dans le modèle SPMD, comment distingue-t-on les différents processus ?

A chaque processus est associé un identificateur de tâche (*tid*). Cet entier permet de repérer le processus qui émet ou reçoit les messages. Des fonctions PVM permettent de connaître cet identificateur (*gettid*, *mytid*). En plus, sur T3D (et sur les premières versions du système de la T3E), à chaque processus est associé un processeur. Cray a implémenté des fonctions PVM supplémentaires permettant de connaître le numéro d'un processeur (*mynode*) ou d'associer le numéro d'un processeur et d'un processus (*getpe*).

6 • Quelles sont les fonctions PVM les plus importantes ?

L'utilisation d'une dizaine de fonctions est, dans la plupart des cas, suffisante pour paralléliser un code. En plus des fonctions d'identification des processus, on distingue les fonctions de communication point à point (*send*, *recv*, *mcast*) et les fonctions de communications collectives (*bcast*, *barrier*, *reduce*, *scatter*, *gather*).

7 • Quels sont les points forts de PVM ?

- Son succès provient surtout de la facilité de portage des codes PVM (malgré quelques ajouts et retraites de fonctions suivant les constructeurs) et de la possibilité d'utiliser une machine parallèle virtuelle (grappe de stations par exemple). Ceci permet de paralléliser son code sur des architectures modestes avant de le passer réellement en exploitation sur des calculateurs parallèles.
- La parallélisation d'un code avec PVM peut-être très efficace.
- PVM est utilisable à partir des langages Fortran, C et C++.

8 • ...et ses points faibles ?

Ce sont les points faibles inhérents à la programmation parallèle par échange explicite de messages, à savoir :

- une certaine lourdeur de programmation,
- des difficultés pour le débogage.

9 • Existe-t-il des règles simples pour minimiser les risques d'erreurs dans l'écriture d'un code PVM ?

Il faut absolument utiliser toutes les fonctionnalités de votre langage de programmation favori. En Fortran, il est conseillé :

- d'utiliser la directive `IMPLICIT NONE` pour déclarer toutes ses variables,
- de dimensionner absolument tous ses tableaux (les `tab(*)` sont à proscrire),
- d'utiliser les nouveautés de Fortran 90 : interface de fonctions, contrôle des bornes de tableaux...

D'autre part, il est toujours plus facile de déboguer sur 2 processeurs que sur 128. Dans la mesure du possible, il faut donc essayer d'avoir un code souple ne dépendant pas d'un nombre fixe de processus.

10 • Existe-t-il d'autres moyens pour paralléliser un code ?

Oui, bien sûr, suivant les architectures il existe des mécanismes de copie mémoire à mémoire (SHMEM sur T3D/T3E) ou des techniques de parallélisme de données (CRAFT et HPF sur T3D/T3E). Mais en ce qui concerne la programmation par passage de message, PVM a su très vite s'imposer pour devenir un standard de fait. Néanmoins, la volonté de définir un véritable standard a conduit des universitaires et des constructeurs à proposer une nouvelle bibliothèque : MPI (Message Passing Interface). Celle-ci offre plus de primitives de communications que PVM. Lors du choix final, MPI est donc à prendre en considération.

J.M. Dupays

POUR EN SAVOIR PLUS

- contactez l'assistance par téléphone, (1) 69 82 42 00, ou bien par courrier électronique, assist@idris.fr
- lisez la note technique numéro 2 (NT2.ps) sur Hebe dans le répertoire `/u/idris/public/notes_techniques`



Du côté de la visualisation

L'IDRIS vient de se doter d'un serveur de visualisation appelé Rhodes. Ce serveur est une Power Challenge XL de chez Silicon Graphics Inc. qui dispose de 40 processeurs R8000 et de 2 gigaoctets de mémoire centrale.

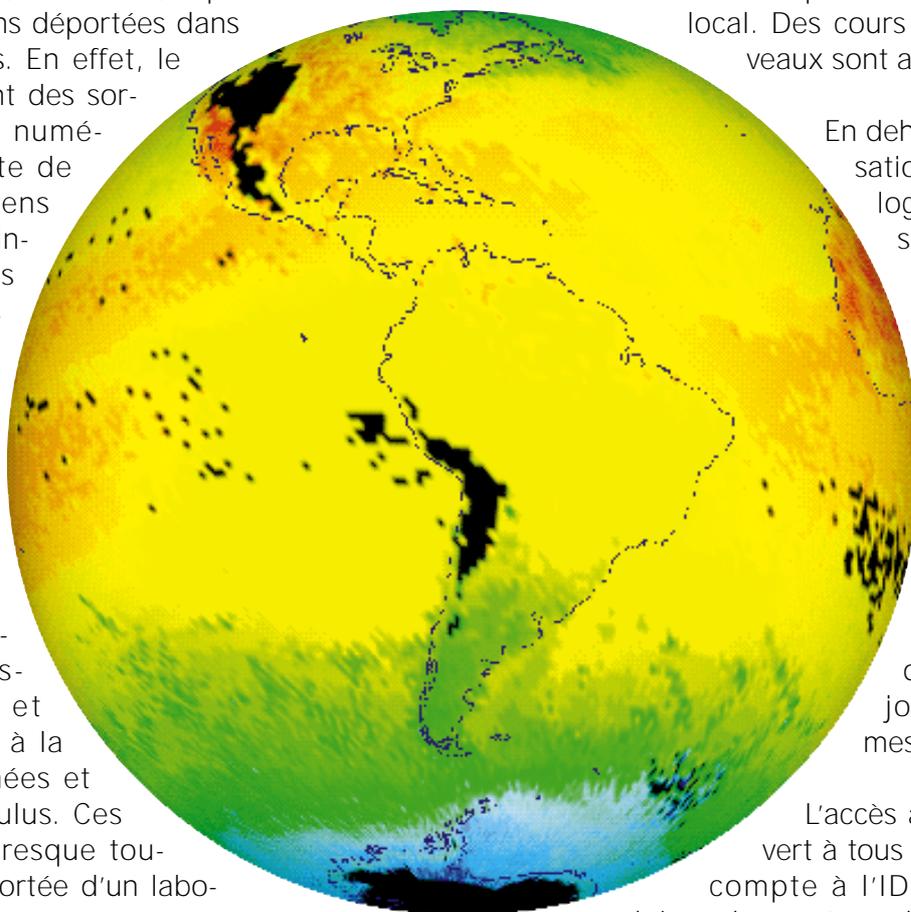
Cette nouvelle machine assure une interface très étroite avec les machines de production de l'IDRIS (parallèle, vectorielle, superscalaire) et les stations déportées dans les laboratoires. En effet, le volume croissant des sorties de calculs numériques nécessite de nouveaux moyens d'analyse et d'interprétation des résultats. Les outils de visualisation sont devenus suffisamment performants pour remplir cette fonction mais ces logiciels requièrent également des ressources CPU et mémoire, liées à la taille des données et traitements voulus. Ces moyens sont presque toujours hors de portée d'un laboratoire.

Face à ce constat, l'IDRIS a mis à la disposition de ses utilisateurs un service national de post-traitement des données via le réseau RENATER.

Les fichiers à analyser restent à l'IDRIS sur Rhodes, qui fournit l'essentiel des ressources nécessaires CPU et mémoire pour le traitement ; puis l'affichage et la manipulation des images s'effectuent sur les stations des utilisateurs. Afin de limiter les latences induites par la charge du

réseau, l'IDRIS préconise l'utilisation de logiciels fonctionnant en réparti entre un serveur central et des stations distantes.

Pour la visualisation 3D interactive, l'IDRIS a opté pour le logiciel AVS utilisé en réparti entre les stations des laboratoires et le serveur de visualisation. Des licences AVS sont disponibles pour les utilisateurs qui n'en disposent pas en local. Des cours de différents niveaux sont assurés.



En dehors de cette utilisation de Rhodes, les logiciels de visualisation 2D-3D présents sur la frontale Hebe et les C90 ont été installés sur le serveur (NCAR et HDF par exemple). Suivant les besoins des utilisateurs, d'autres produits seront ajoutés au fur et à mesure.

L'accès à Rhodes est ouvert à tous les titulaires d'un compte à l'IDRIS. Les formulaires d'ouverture de compte et les modalités de fonctionnement sont disponibles sur le serveur Web de l'IDRIS (<http://www.idris.fr>) ou sur la frontale Hebe sous la forme d'un fichier PostScript *graph.ps* sous */u/idris/public/formulaires*.

Pour tout renseignement complémentaire et/ou pratique, vous pouvez prendre contact avec le groupe Visualisation & Vidéo de l'IDRIS à graph@idris.fr.





Formation IDRIS

Calendrier des cours IDRIS pjusqu'au 30 juin 1996 :

13 mai	Aspect d'UNIX : administration et sécurité	2 jours
21 mai	Fortran 90 (organisé à Toulouse)	3 jours
23 mai	Chimie : utilisation du logiciel Gaussian	2 jours
3 juin	Cray C90 : architecture vectorielle, utilisation du C98 et C94	5 jours
6 juin	Cray T3D : introduction et utilisation de PVM (organisé à Toulouse)	2 jours
10 juin	Cray T3D : introduction et utilisation de PVM	2 jours
12 juin	Cray T3D : introduction, CRAFT, optimisation, outils interactifs	3 jours
17 juin	Cray T3D : introduction de MPI	3 jours
17 juin	Le langage C	5 jours
24 juin	Applications scientifiques de C++	2 jours

L'IDRIS assure également des cours en province. Les demandes doivent être formulées à dir@idris.fr. Pour toute inscription ou tout renseignement complémentaire, veuillez contacter le secrétariat de l'IDRIS au (1) 69 82 41 46/41 67 ou par messagerie au secretariat@idris.fr.



Numéros de téléphone

Attention, l'IDRIS se dote d'un nouvel autocommutateur à compter du 1er juin 1996. Notez dès à présent les nouveaux numéros :

Direction	(1) 69 35 85 85
Secrétariat direction	(1) 69 35 85 05
Assistance	(1) 69 35 85 55
Télécopie	(1) 69 85 37 75



Renseignements pratiques

Qui joindre à l'IDRIS :

La Direction	(1) 69 82 41 41 et dir@idris.fr
Le Secrétariat	(1) 69 82 41 77/41 46/41 67 et secretariat@idris.fr
L'Assistance	(1) 69 82 42 00 et assist@idris.fr
Le Pupitre	(1) 69 82 41 50
Par FAX	(1) 69 28 52 73



Serveur Web de l'IDRIS

L'IDRIS a ouvert son serveur Web accessible à l'adresse suivante : <http://www.idris.fr>

Vous y trouverez des informations générales concernant l'IDRIS, des FAQ sur l'utilisation des machines du centre, des documentations et des cours, le différents formulaires d'inscription.



Documentation et Publications IDRIS

Liste des notes techniques disponibles ou en cours de rédaction :

NT 15	Introduction à Mopac dans l'environnement de l'IDRIS
NT 16	Utilisation de Gamess à l'IDRIS
NT 17	Introduction au serveur de visualisation de l'IDRIS



Nouvelles publications IDRIS disponibles

Utilisation de MPI en décomposition de domaine (I. Brugeas)
Introduction à AVS (T. Goldmann)
AVS avancé (T. Goldmann)

CNRS CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

idris

Directeur de la Publication

Victor Alessandrini

Rédacteur en chef

Thierry Goldmann

Merci de photocopier et de renvoyer cette demande à :
IDRIS - Bat. 506 - B.P. 167 - 91403 ORSAY CEDEX - FRANCE

Je souhaite recevoir La Lettre de l'IDRIS J'ai déjà un login à l'IDRIS Je n'ai pas de login à l'IDRIS

Nom : Prénom : Fonction :

Organisme :

Adresse :

Code Postal : Ville : Pays :

Tél : Fax :