Hierarchical Components for the GRID

Denis Caromel, et al.
ProActive.ObjectWeb.org
OASIS Team
INRIA -- CNRS - I3S -- Univ. of Nice Sophia-Antipolis, IUF
Santiago, Nov. 9 2004

0. GRIDs
1. *ProActive:* Asynchronous Distributed Objects
2. Groups
3. Components

ProActive:
A Java API + Tools for Parallel, Distributed Computing

- A uniform framework: An Active Object pattern
- A formal model behind: Prop. Determinism, insensitivity to deploy.

Programming Model:

- **Remote Objects (Classes, not only Interfaces, Dynamic)**
- **Asynchronous Communications, Futures, Wait-By-Necessity**
- **Groups, Mobility, Components, Security, Fault Tolerance: Checkpoints**

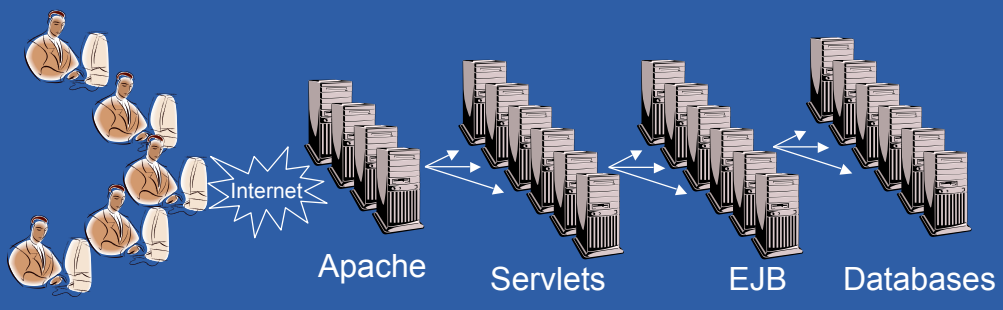
Environment:

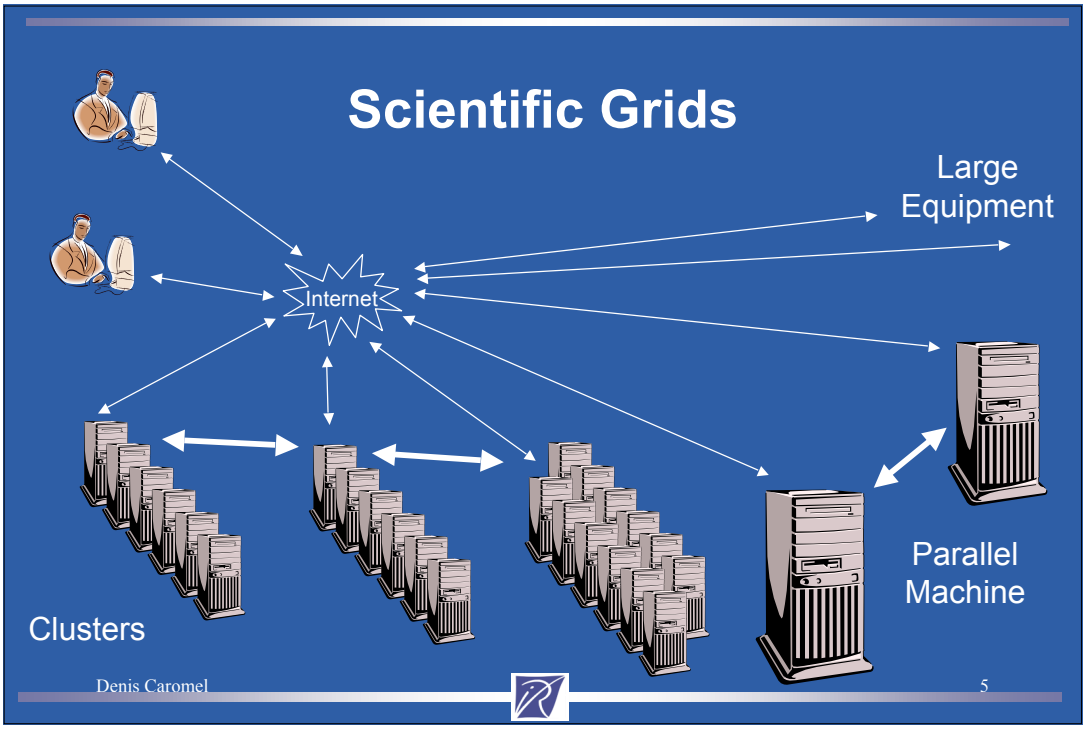
- **XML Deployment Descriptors,** `Web Service Export., HTTP, ssh Tunneling`
- **Various protocols:** `rsh,ssh,LSF,PBS,Globus,Sun Grid Engine, sshGSI`
- **Visualization and monitoring:** IC2D

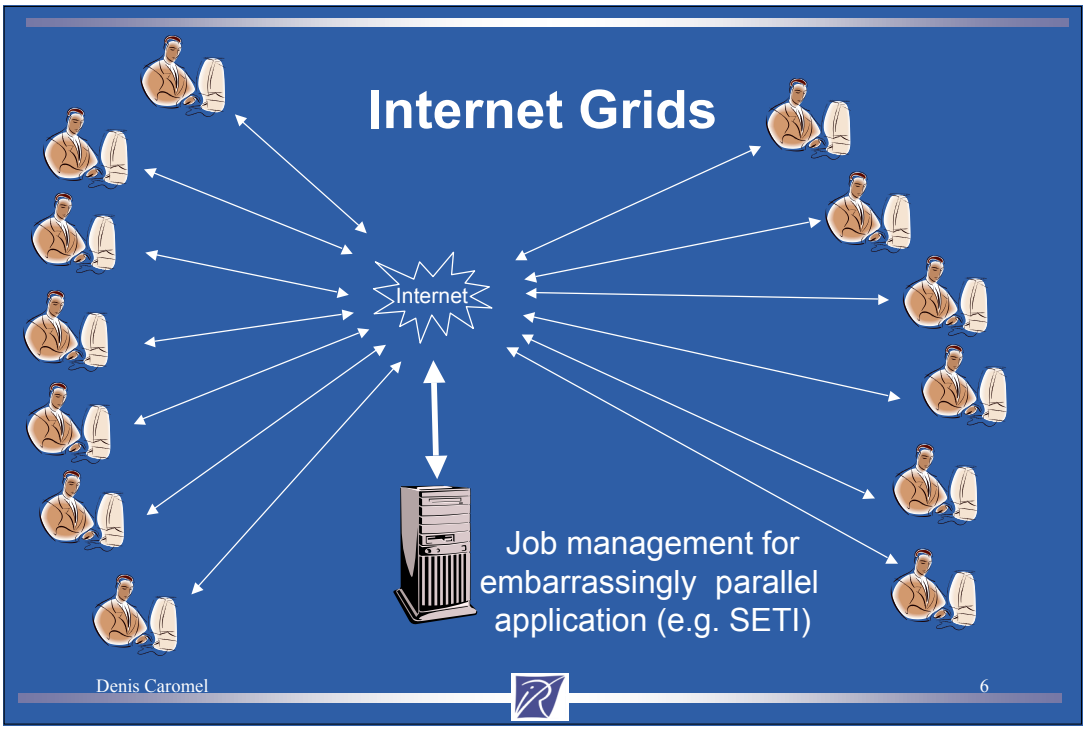
In the www. ObjectWeb .org Consortium (Open Source middleware)
since April 2002 (LGPL license)

GRIDs

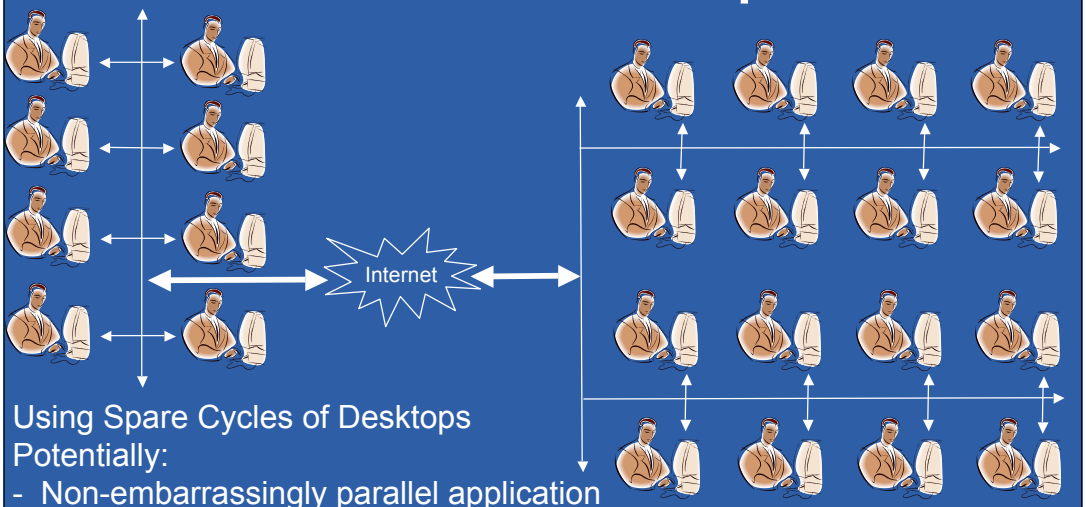
Enterprise Grids




Scientific Grids
Large
Equipment
Internet
Clusters
Parallel
Machine
Denis Caromel
5


Internet Grids
Internet
Job management for
embarrassingly parallel
application (e.g. SETI)
Denis Caromel
6

Intranet Grids - Desktop Grids



Using Spare Cycles of Desktops
Potentially:
- Non-embarrassingly parallel application
- Across several sites

The multiple GRIDs

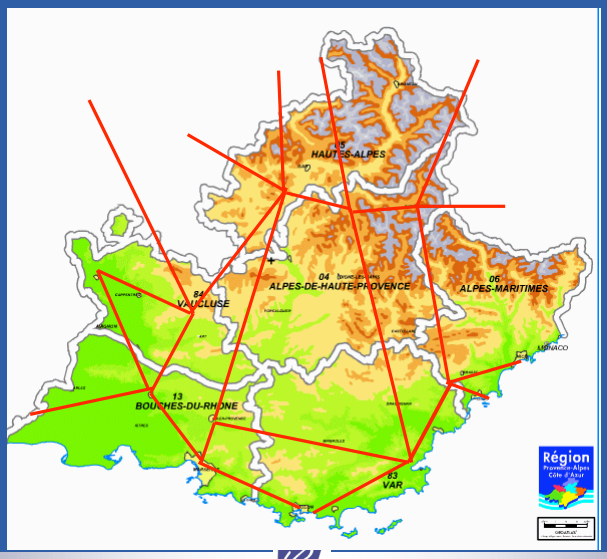
- Scientific Grids
- Enterprise Grids
- Internet Grids
- Intranet Desktop Grids

Strong convergence in process!

At least at the infrastructure level, i.e. WS

Grid: from enterprise ... to regional

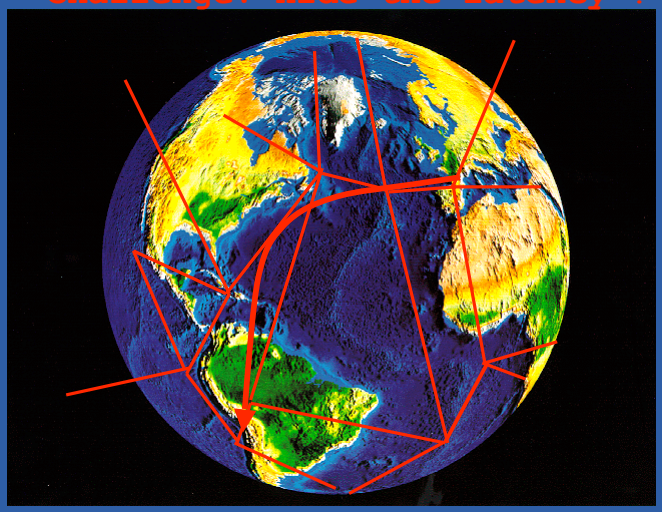
Very hard deployment problems … right from the beginning



Grid: from regional ... to worldwide

Communication Nice - Santiago: 70 ms Light Speed

Challenge: Hide the latency !



Define adequate programming model

Programming

W r a p p i n g

Composing

Deploying

Figures: Web Page Hits: ~ 3 000 / month,
Downloads: 150-300 / month, Users: ?? .us, .mx, .br, .cl, .ch, .it, .fr, ...

Distributed Objects

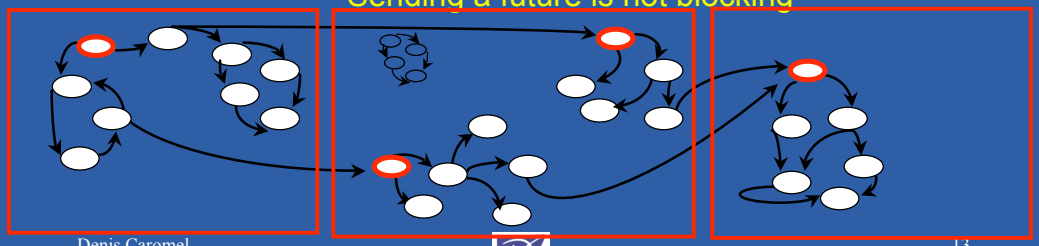
ProActive Programming

ProActive model

Java RMI (Remote Method Invocation = Object RPC = o.foo(p))

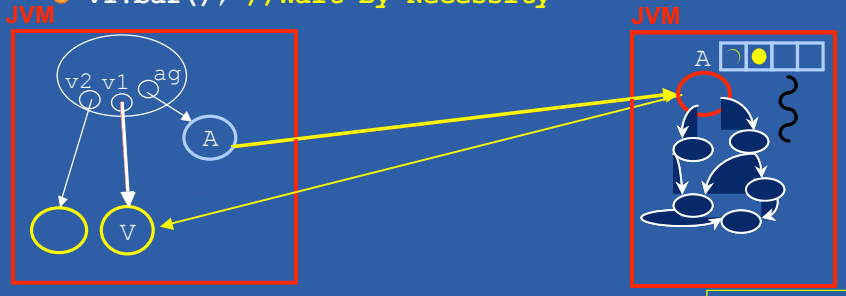
plus a few important features:

- Asynchronous Method calls towards Active Objects:
  Implicit Futures as RMI results
- Wait-By-Necessity:
  - Automatic wait upon the use of an implicit future
  - First-Class Futures:
    - Futures passed to other activities
    - Sending a future is not blocking



ProActive : Active objects

```
A ag = newActive ("A", […], VirtualNode)
V v1 = ag.foo (param);
V v2 = ag.bar (param);
...
v1.bar(); //Wait-By-Necessity
```

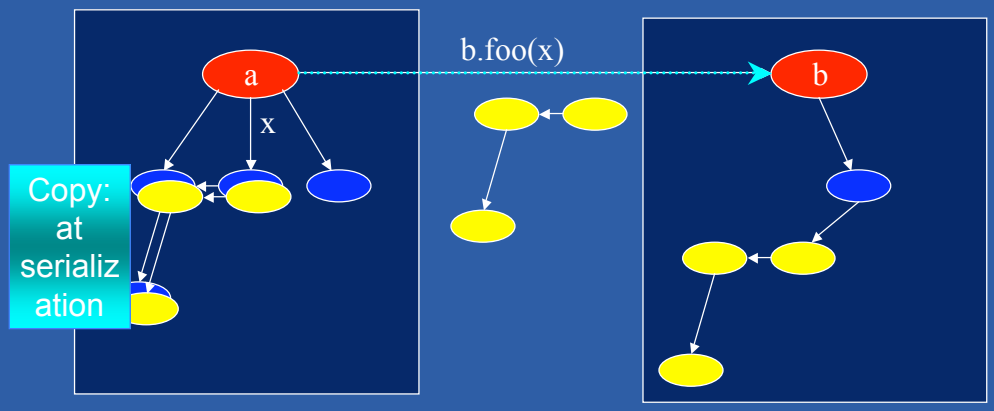


Java Object Active Object Req. Queue

Future Object Proxy Request Thread

Wait-By-Necessity is a Dataflow Synchronization

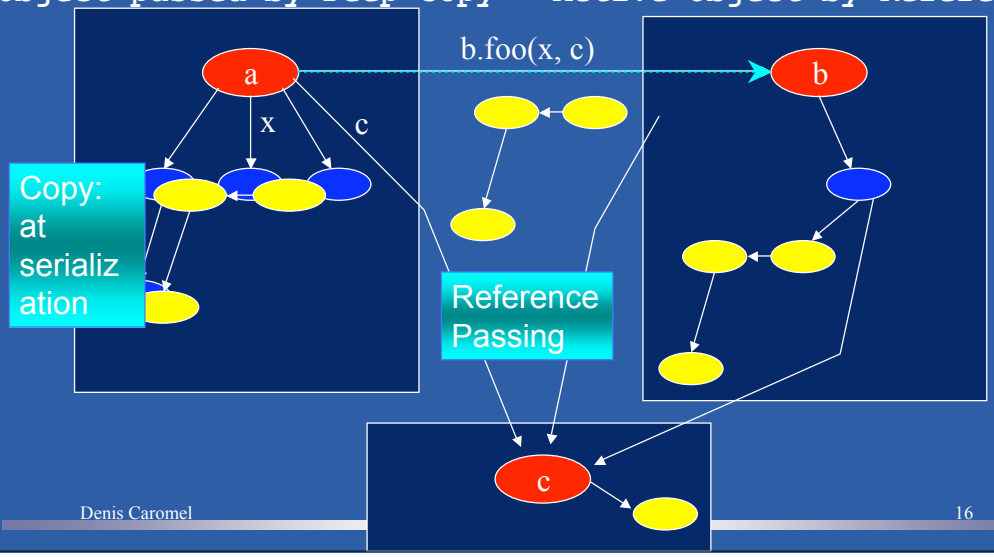
Call between Objects: Parameter passing: Copy of Java Objects



(Deep) Copies evolve independently -- No consistency

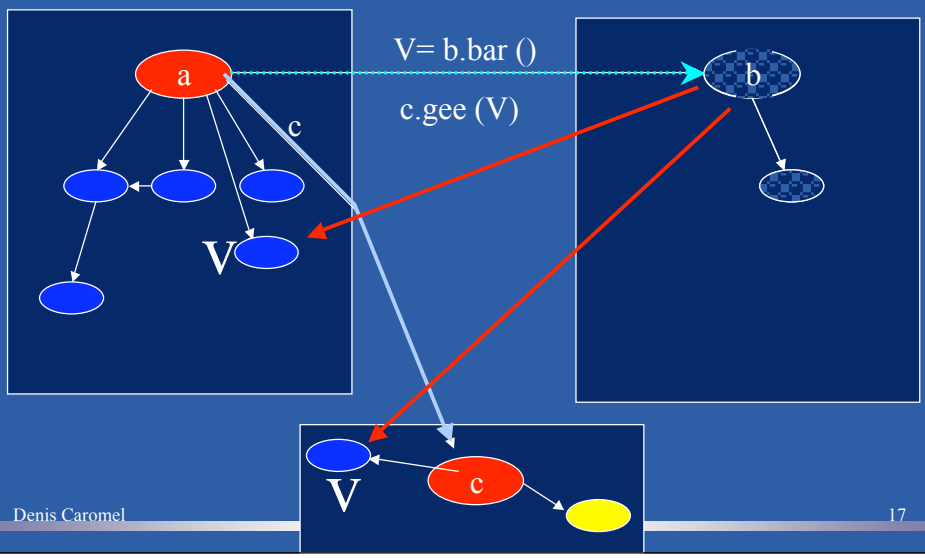
Call between Objects: Parameter Passing: Active Objects

Object passed by Deep Copy - Active Object by Reference



Wait-By-Necessity: First Class Futures

Futures are Global Single-Assignment Variables



ProActive : Explicit Synchronizations

```
A ag = newActive ("A", […], VirtualNode)
V v = ag.foo(param);
...
v.bar(); // Wait-by-necessity
```

Single Future Synchronization:

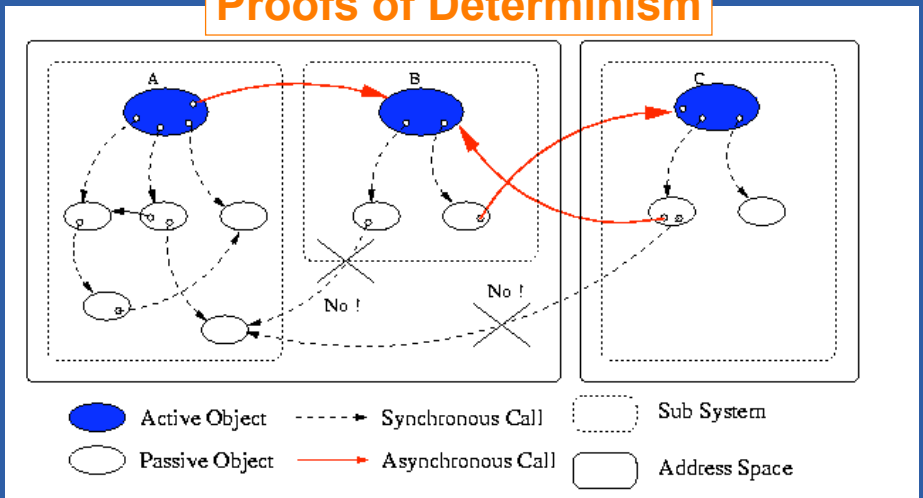
- `ProActive.isAwaited (v); // Test if available`
- `.waitFor (v); // Wait if not available`

Vectors of Futures:

- `.waitForAll (Vector); // Wait all of them`
- `.waitForAny (Vector); // Get One`

Standard system at Runtime: Asynchrony, WbN, ... but no sharing

Proofs of Determinism



Groups

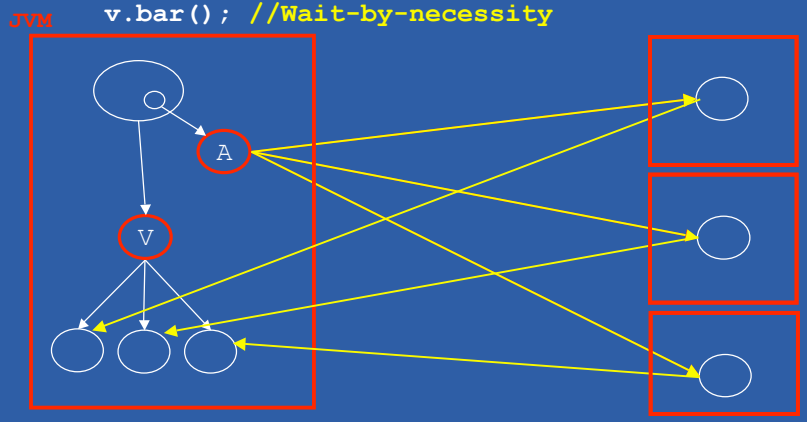
Collective Communications: Groups

- Manipulate groups of Active Objects, in a simple and typed manner:
  - ⇒ Typed and polymorphic Groups of active and remote objects
  - ⇒ Dynamic generation of group of results
  - ⇒ Language centric, Dot notation
- Be able to express high-level collective communications (like in MPI):
  - broadcast,
  - scatter, gather,
  - all to all

```
A ag=(A)ProActiveGroup.newActiveGroup(«A»,{{p1},...},{Nodes,..});
V v = ag.foo(param);
v.bar();
```


Creating AO and Groups

```
A ag = newActiveGroup ("A", […], VirtualNode)
V v = ag.foo(param);
...
v.bar(); //Wait-by-necessity
```



Object-Oriented
Typed Group Communications

Group, Type, and Asynchrony
are crucial for Cpt. and GRID

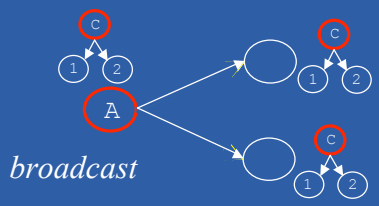
Broadcast or Scatter

Broadcast is the default behavior

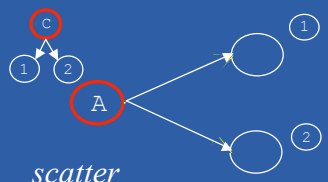
Scatter is also possible

- use a group as parameter
- Scattered depends on rankings

```
gA.bar(gC);    // broadcast gC
ProActive.setScatterGroup(gC);
ga.bar(gC);    // scatter gC
```



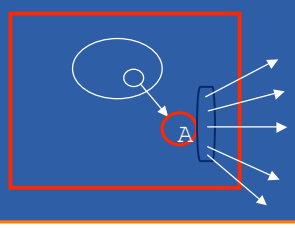
broadcast

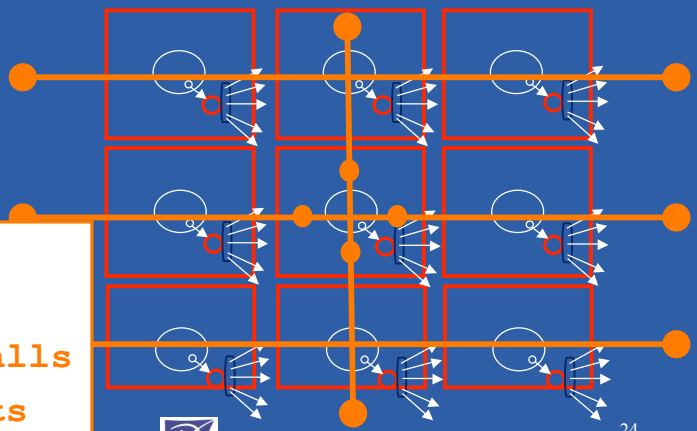


scatter

OO SPMD

```
A ag = newSPMDGroup ("A", […], VirtualNode)
        // In each member
        myGroup.barrier ("2D"); // Global Barrier
        myGroup.barrier ("vertical"); // Any Barrier
        myGroup.barrier ("north","south","east","west");
```





Still,

not based on raw messages, but

Typed Method Calls

==> Components

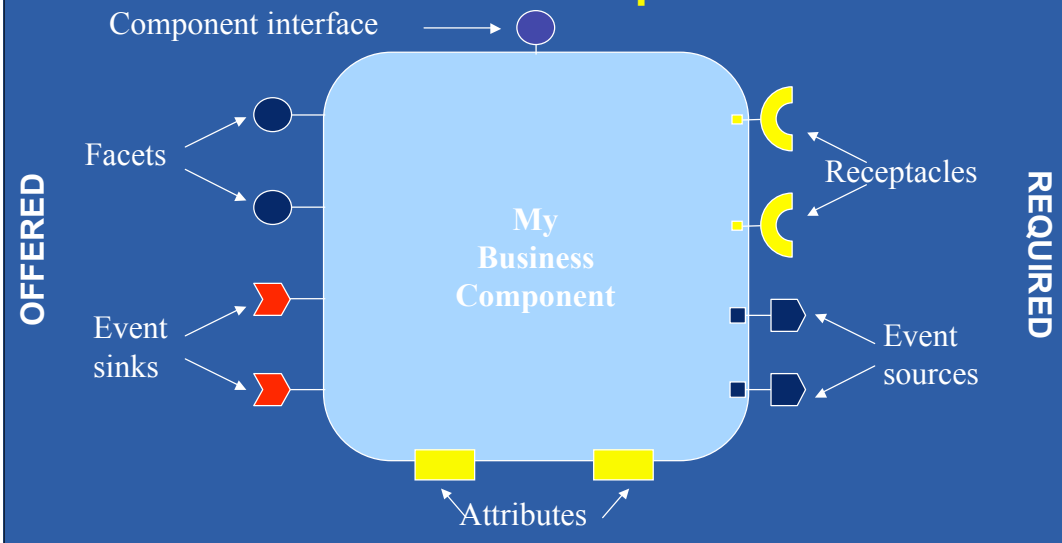
Parallel, Distributed, Hierarchical

Components

for the Grid

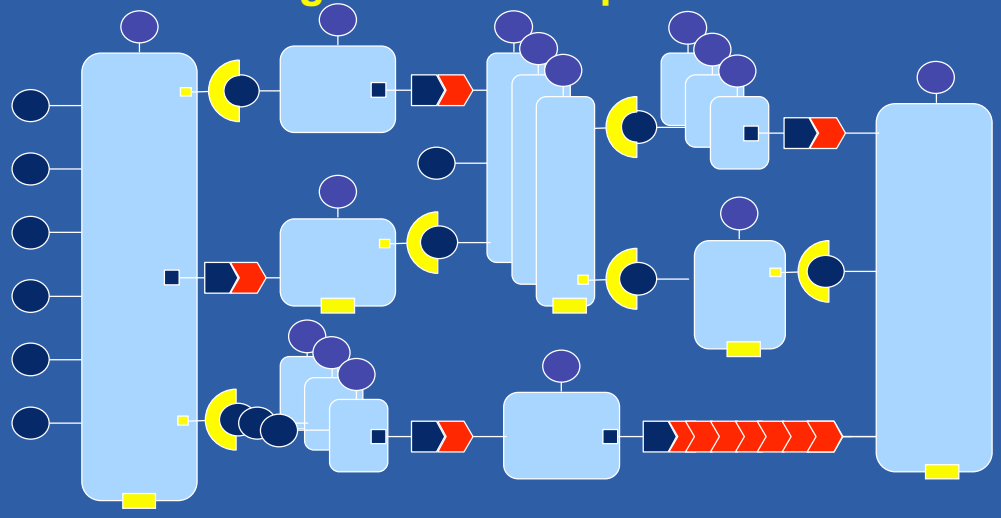
Composing

A CORBA Component



Courtesy of Philippe Merle, Lille, OpenCCM platform

Building CCM Applications = Assembling CORBA Component Instances



Provide + Use, but flat assembly

Component Orientedness

- Level 1: Instantiate - Deploy - Configure
  - Simple Pattern
  - Meta-information (file, XML, etc.) — JavaBeans, EJB
- Level 2: Assembly (flat)
  - Server and client interfaces — CCM
- Level 3: Hierarchic
  - Composite — Fractal, ProActive, ...
- Level 4: Distributed + Reconfiguration
  - Binding, Inclusion, Location — ProActive + On going work

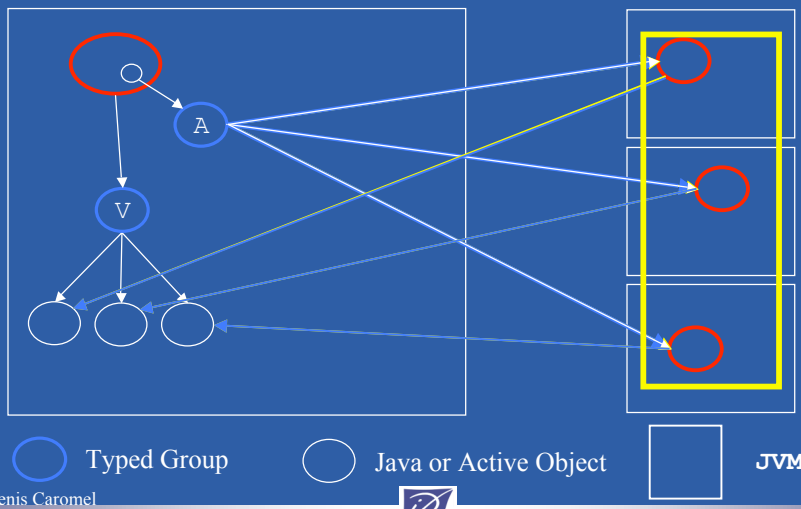
Interactions / Communications: `ProActive`

Functional Calls: service, event, stream

Non-Functional: instantiate, deploy, start/stop, inner/outer, re-bind

Distributed Components (1)

```
ComponentIdentity Cpt = newActiveComponent (params);
A a = Cpt … .getFcInterface ("interfaceName");
V v = a.foo(param);
```

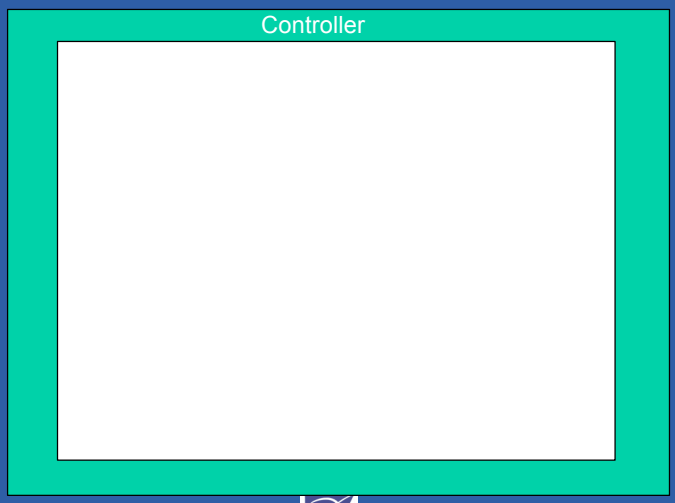


Example of component instance

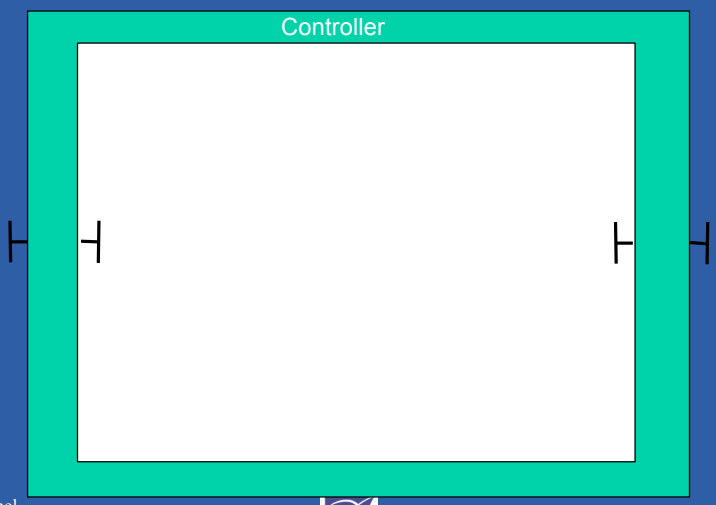
Typed Group Java or Active Object JVM

The Fractal model: Hierarchical Component

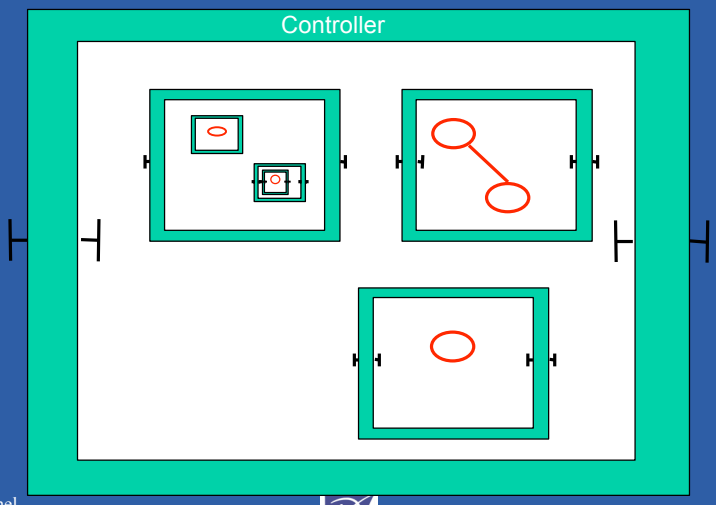
Defined by E. Bruneton, T. Coupaye, J.B. Stefani, FT, et al.



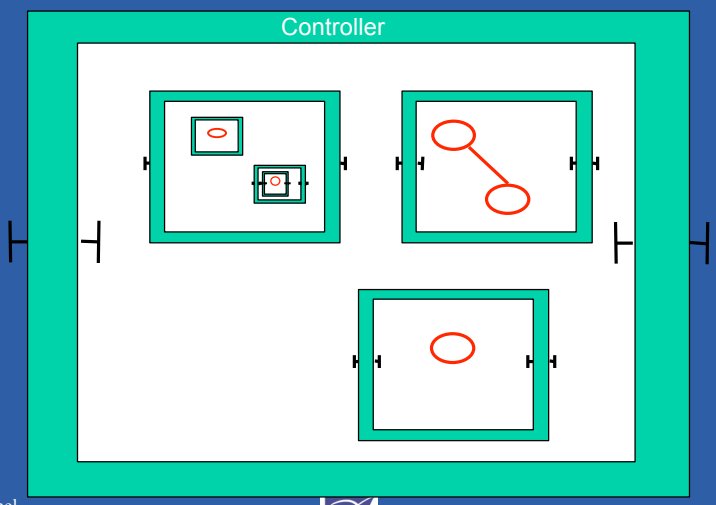
Interface = access point



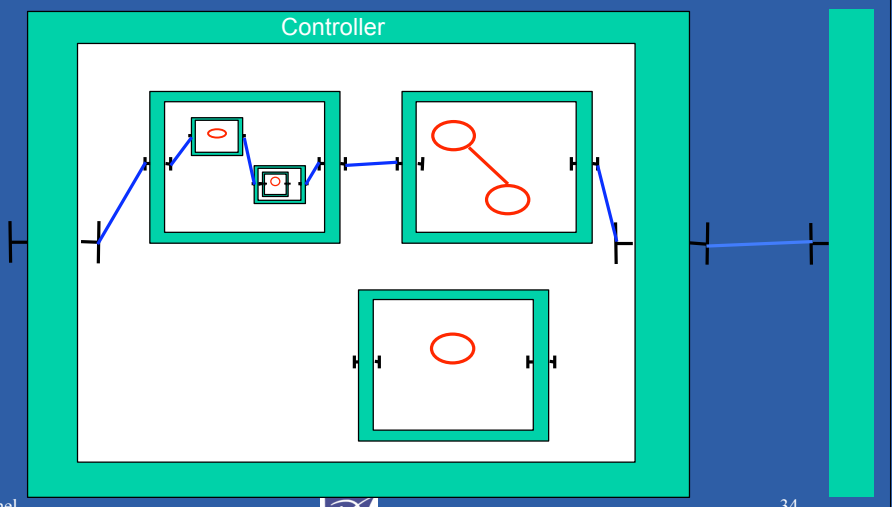
Hierarchical model : composites encapsulate primitives encapsulate Java code



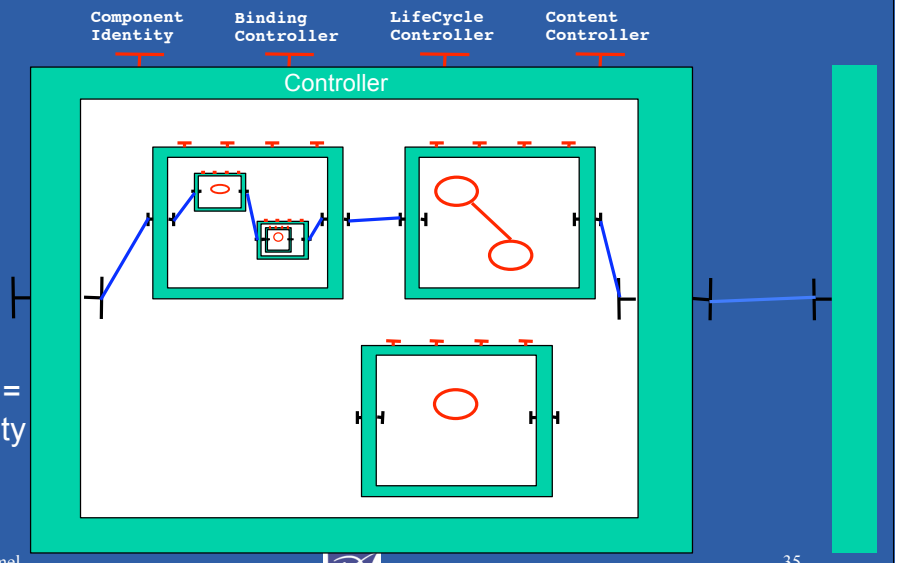
Binding = interaction



Binding = interaction



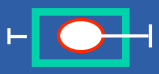
Controllers : non-functional properties



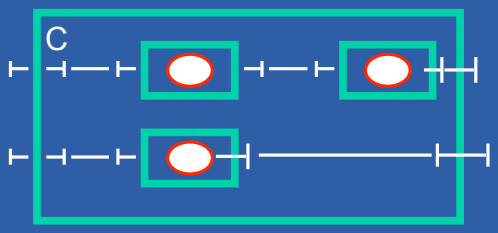
Component = runtime entity

ProActive Components for the GRID

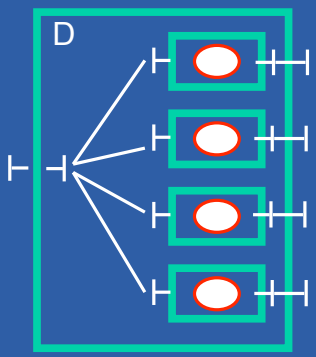
An activity, a process, … potentially in its own JVM



1. Primitive component



2. Composite component



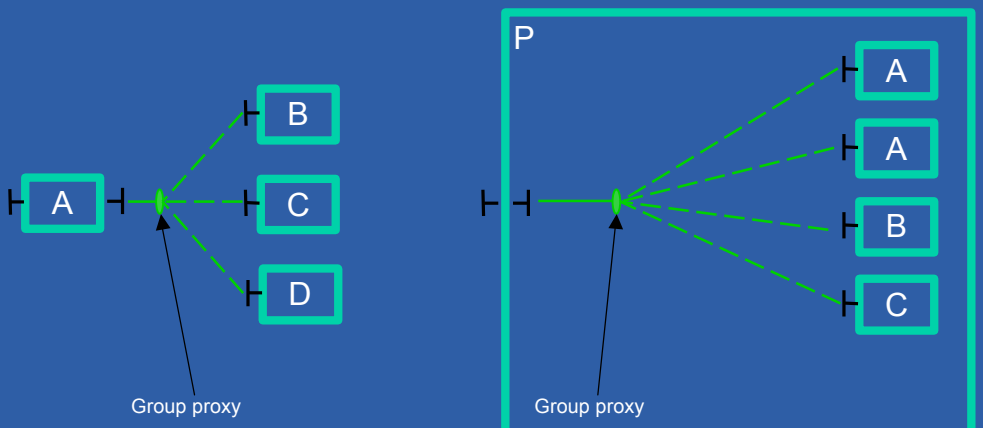
3. Parallel and composite component

Composite: Hierarchical, and Distributed over machines

Parallel: Composite + Broadcast (group)

Groups in Components (1)

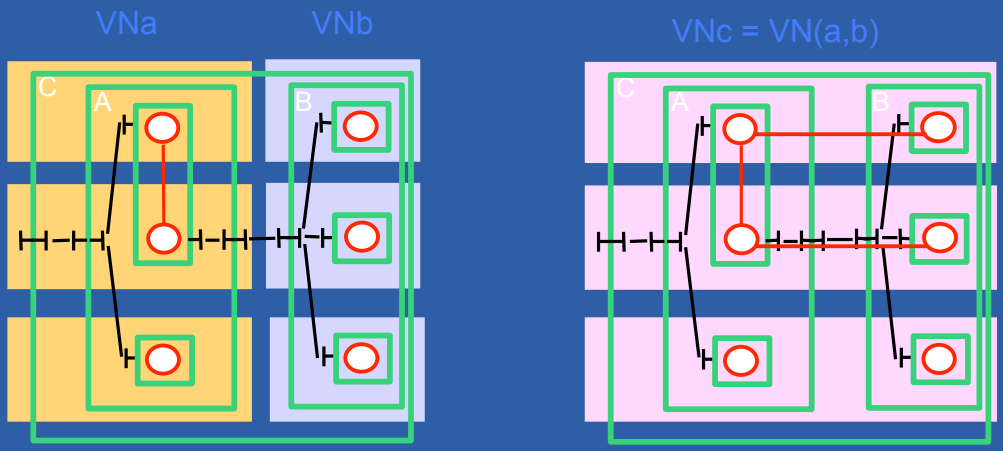
A parallel component!



Broadcast at binding, on client interface

At composition, on composite inner server interface

XML Deployment (Not in source)



Separate or Co-allocation

ProActive Component Definition

A component is:

- Formed from one (or several) Active Object
- Executing on one (or several) JVM
- Provides a set of server ports: Java Interfaces
- Uses a set of client ports: Java Attributes
- Point-to-point or Group communication between components

XML Example

Hierarchical:

- Primitive component: define with Java code and a descriptor
- Composite component: composition of primitive + composite
- Parallel component: multicast of calls in composites

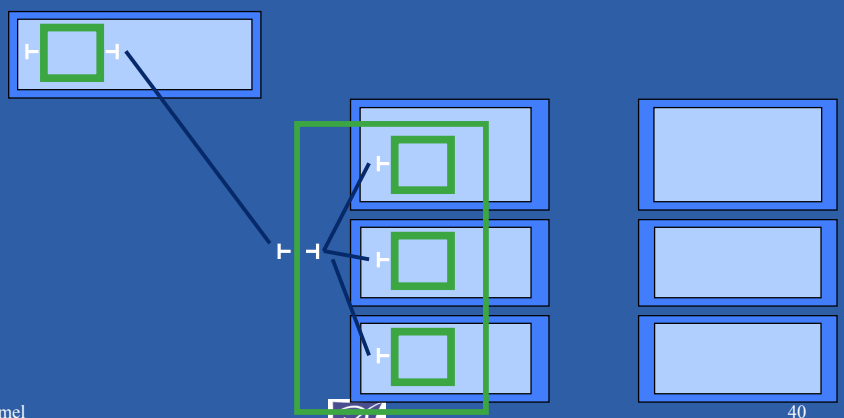
Descriptor:

- XML definition of primitive and composite (ADL)
- Virtual nodes capture the deployment capacities and needs

Virtual Node is a very important abstraction for GRID components

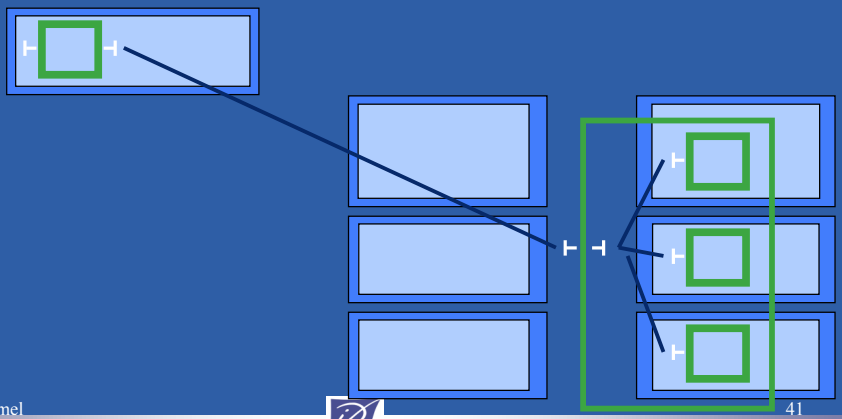
Migration Capability of composites

Migrate sets of components, including composites



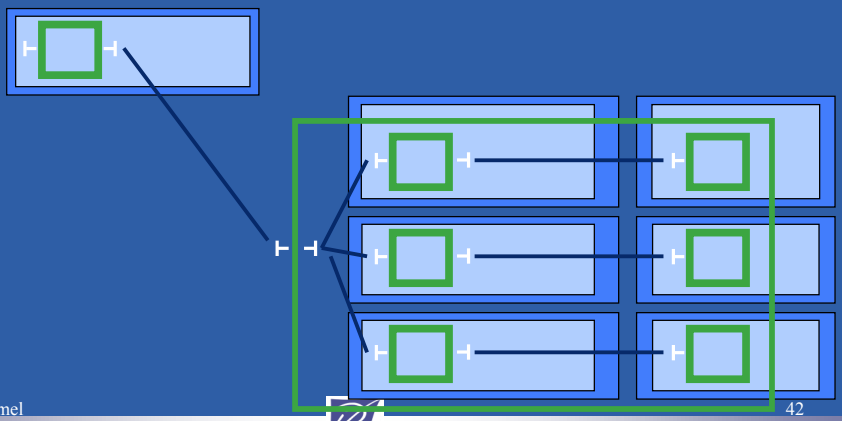
Migration Capability of composites

Migrate sets of components, including composites



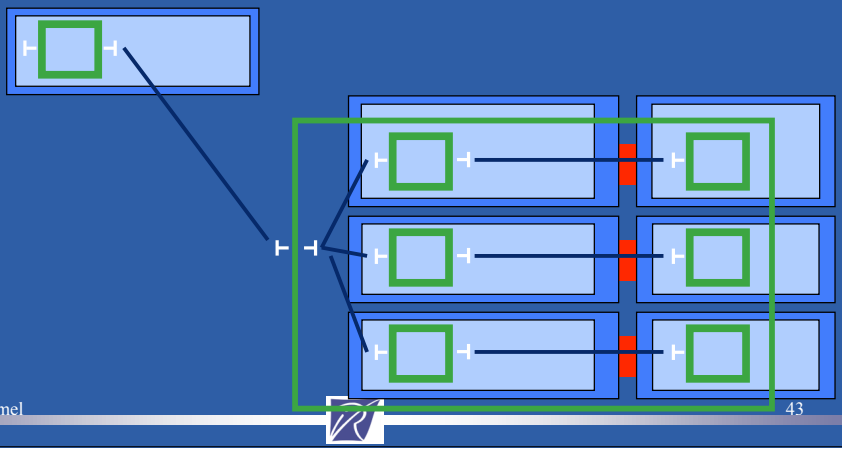
Co-allocation, Re-distribution

e.g. upon communication intensive phase



Co-allocation, Re-distribution

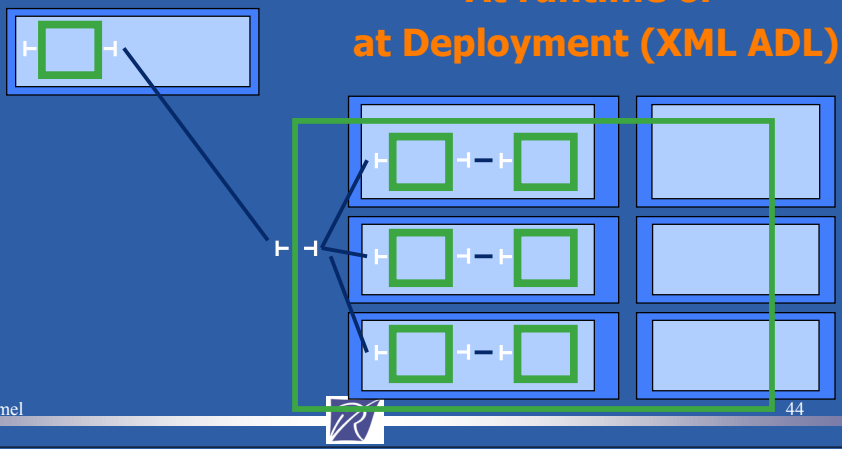
e.g. upon communication intensive phase



Co-allocation, Re-distribution

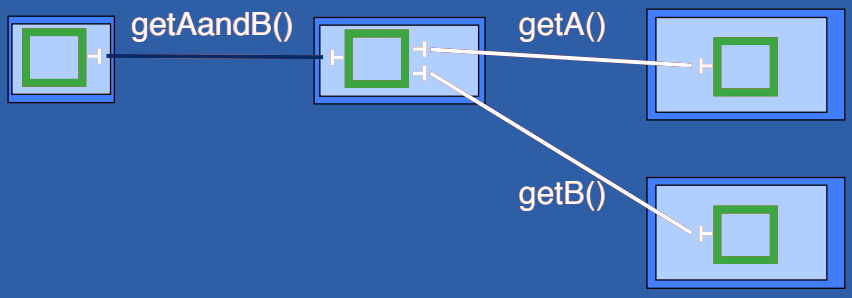
e.g. upon communication intensive phase

At runtime or
at Deployment (XML ADL)



Functionalities : Without First Class Futures

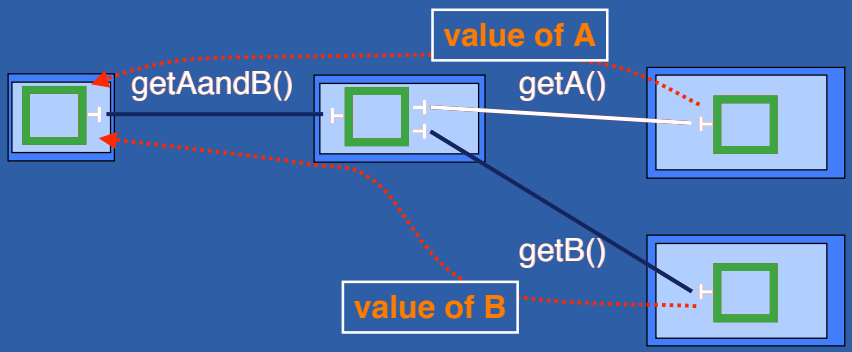
Or in the case of **Synchronous** method calls



Functionalities : With First Class Futures

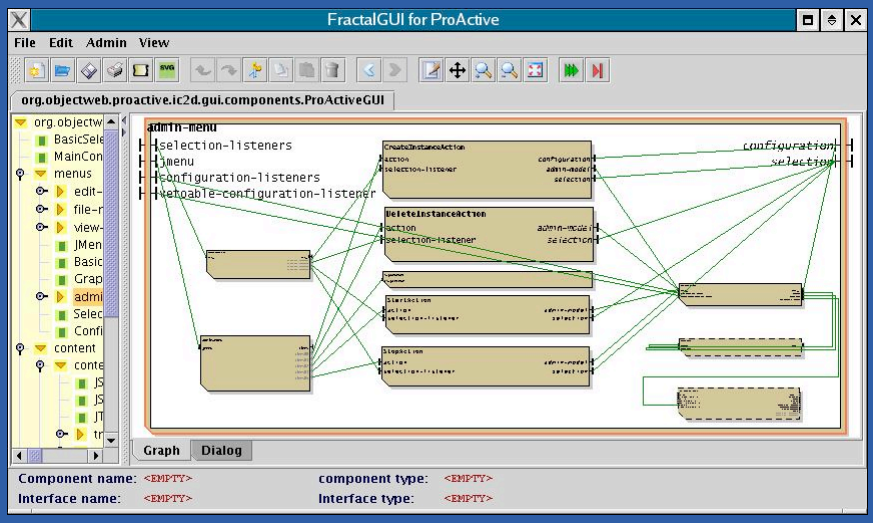
Non-blocking method calls

Example 2 : **Asynchronous** method calls with full-fledge **Wait-By-Necessity**

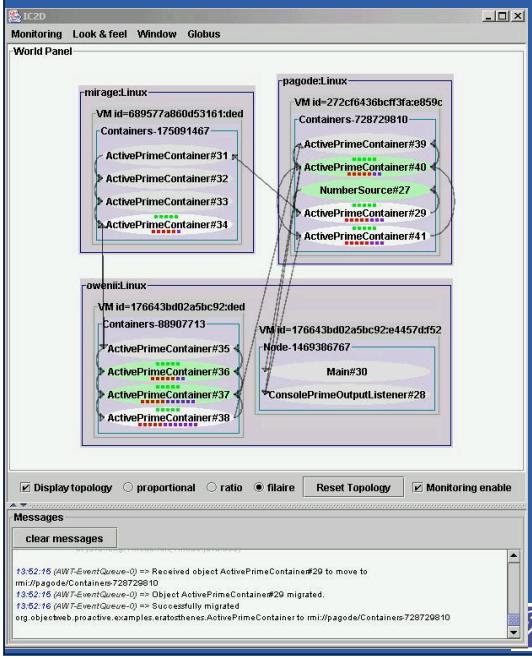


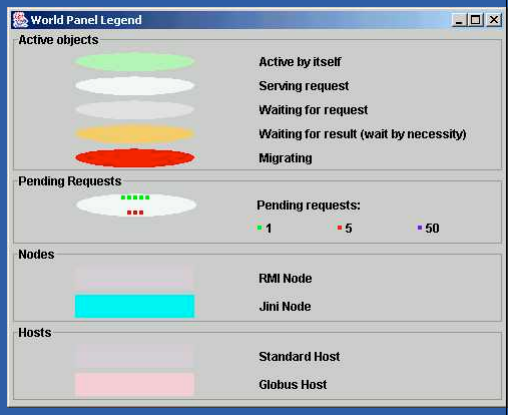
Assemblage are not blocked with Asynchrony + WbN

On-going work : GUI



IC2D: Interactive Control and Debugging of Distribution





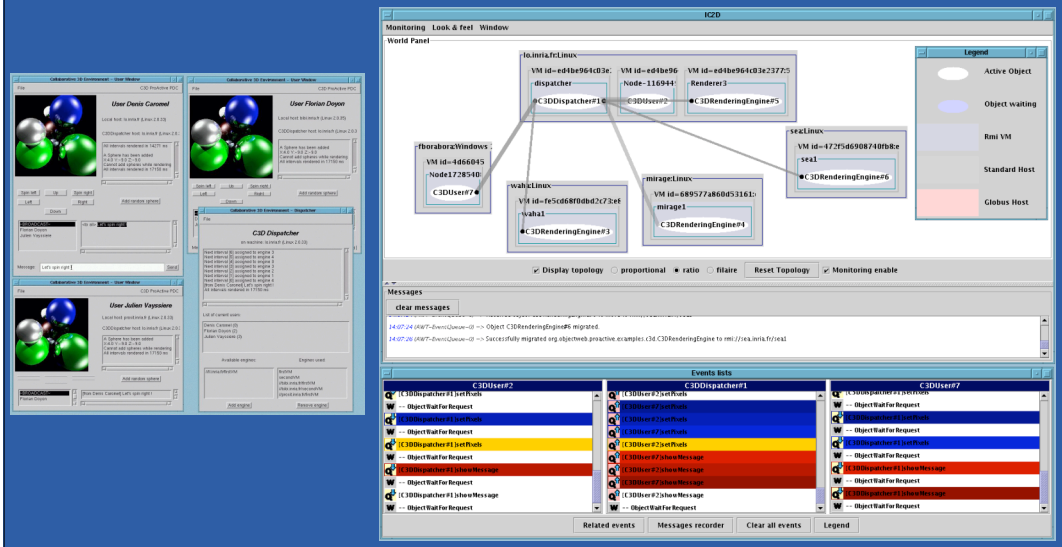
With any ProActive application

Features:

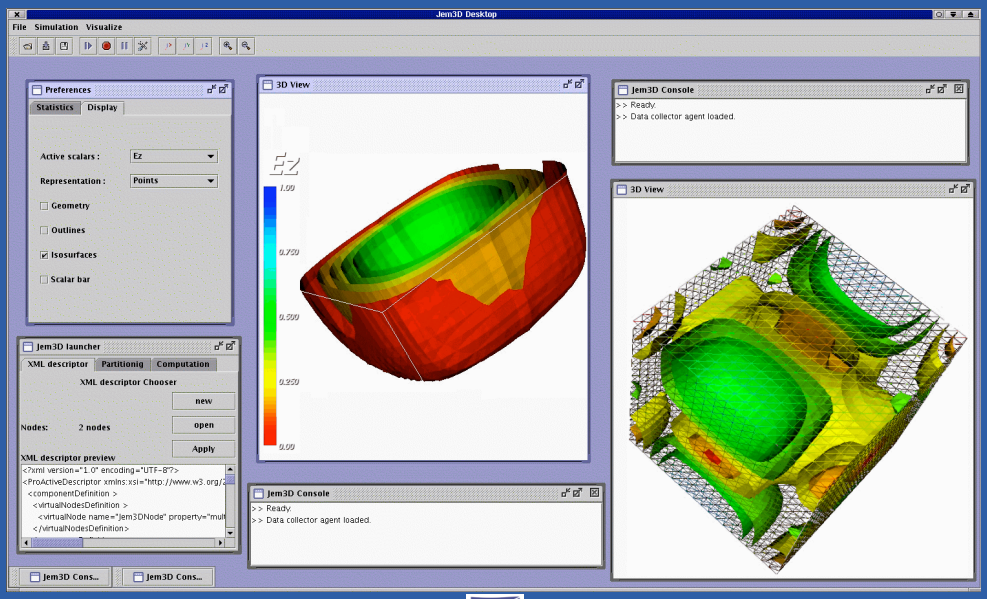
Graphical and Textual visualization

Monitoring and Control

C3D Monitoring: graphical and textual com.



Jem3D



JEM 3D : Java 3D Electromagnetism

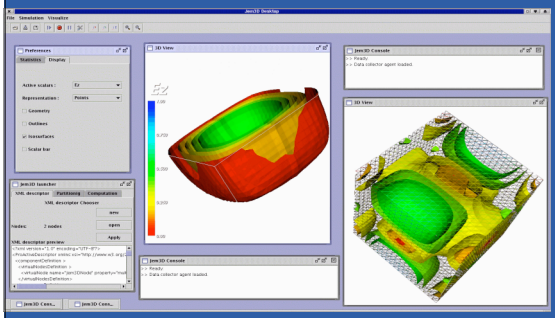
together with Said El Kasmi, Stéphane Lanteri (caiman)

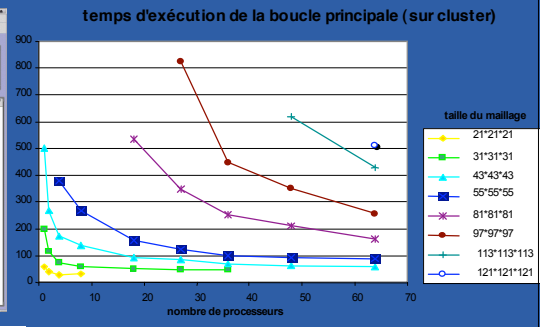
Maxwell 3D equation solver, Finite Volume Method (FVM)

Pre-existing Fortran MPI version: EM3D (CAIMAN team @ INRIA)

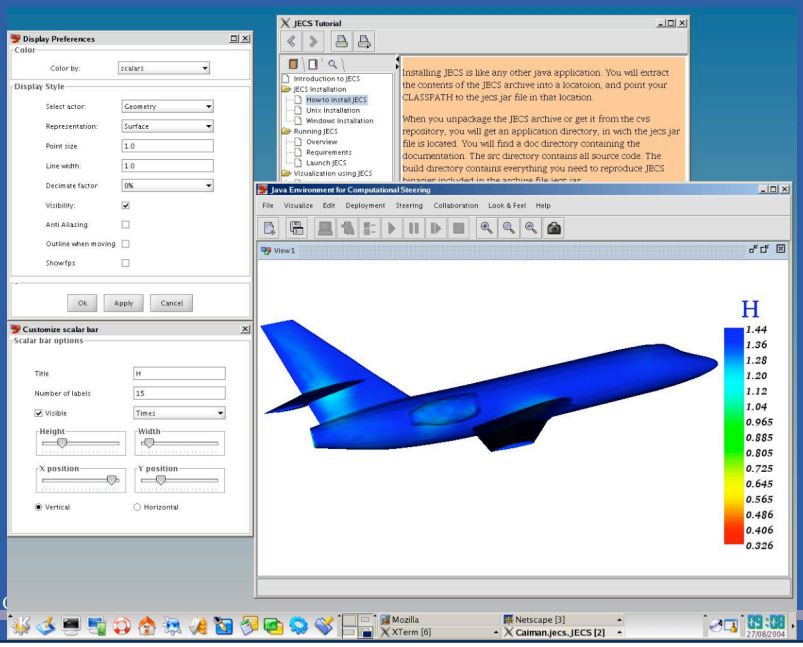
Up to 294 machines at the same time (Intranet and cluster)

Large data sets: 150x150x150 (100 million facets)

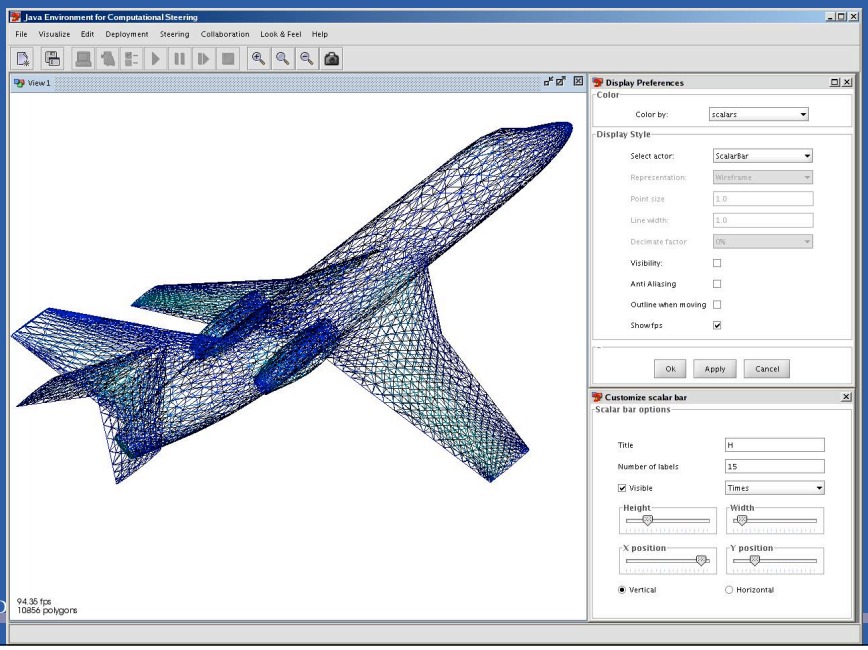




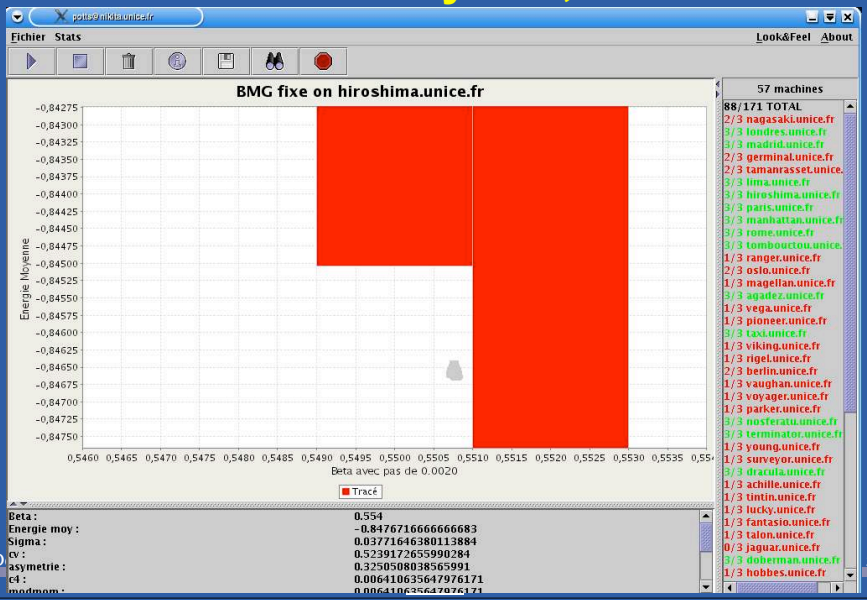
JECS : A Generic Version of Jem3D



JECS : A Generic Version of Jem3D



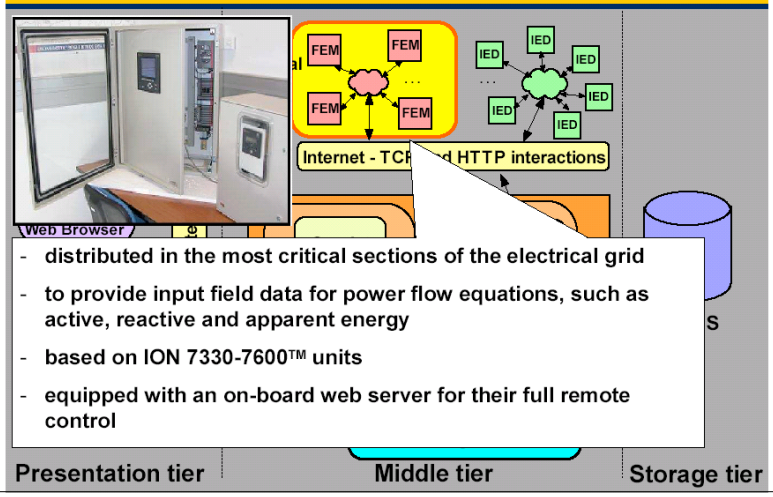
Monte Carlo Simulations, Non-Linear Physics, INLN



Electric Network Planning, E. Zimeo et al., Benevento (Naples), Italy

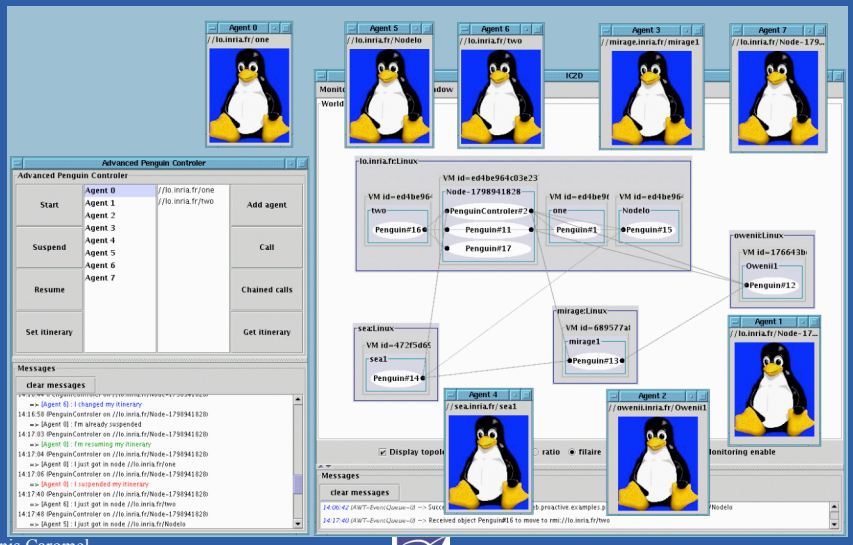
On-line Power Systems Security Analysis (OPSSA)

A network of field power meters (FEMs)

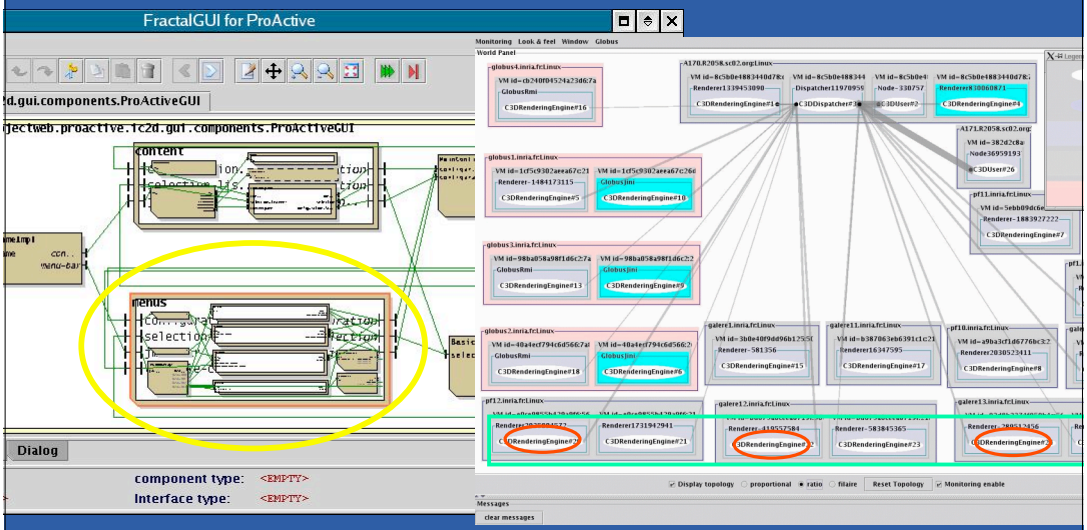


- distributed in the most critical sections of the electrical grid
- to provide input field data for power flow equations, such as active, reactive and apparent energy
- based on ION 7330-7600™ units
- equipped with an on-board web server for their full remote control

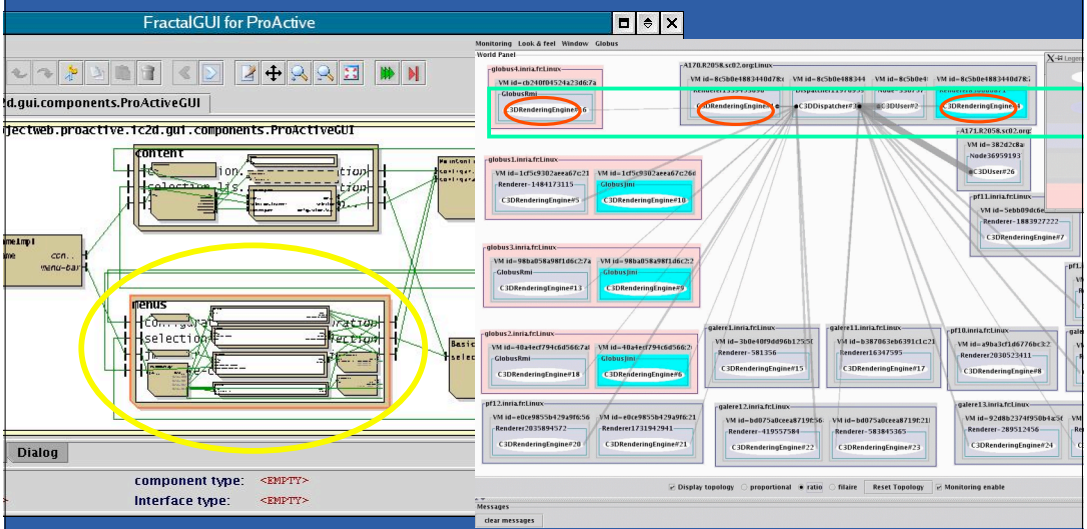
Mobile Application executing on 7 JVMs



Perspective for Components - PSE
Graphical Composition, Monitoring, Migration



Perspective for Components - PSE
Graphical Composition, Monitoring, Migration



Conclusions and A Few Directions

ProActive
Programming, Composing, Deploying on the Grid

A Strong Programming Model + Components

FACTS AND FIGURES

5 years of computation in 17 days in Desktop P2P

Deployed at once on 600 CPUs (Plugtests on ssh, Globus, LSF, ...)

Conclusions and A Few Directions

ProActive
Programming, Composing, Deploying on the Grid

A Strong Programming Model + Components

FACTS AND FIGURES

5 years of computation in 17 days in Desktop P2P

Deployed at once on 600 CPUs (Plugtests on ssh, Globus, LSF, ...)

(Close to) Beating Fortran on an Electromagnetic Application

PERSPECTIVES FOR COMPONENTS

Safe Reconfiguration

How to specify for components: QoS, Ranking, etc. ?

A great alchemy for the Grid:

Asynchrony + Wait By Necessity + Groups + Components

Conclusion - Beating Fortran ?

Current status:

- Sequential Java vs. Fortran code: 2 times slower
- Large data sets in Java ProActive: 150x150x150 (100 million facets)
- Large number of machines: up to 294 machines in Desktop P2P
- Speed up on `16 machines:`
  - Fortran: 13.8
  - ProActive/Ibis: 12
  - ProActive/RMI: 8.8

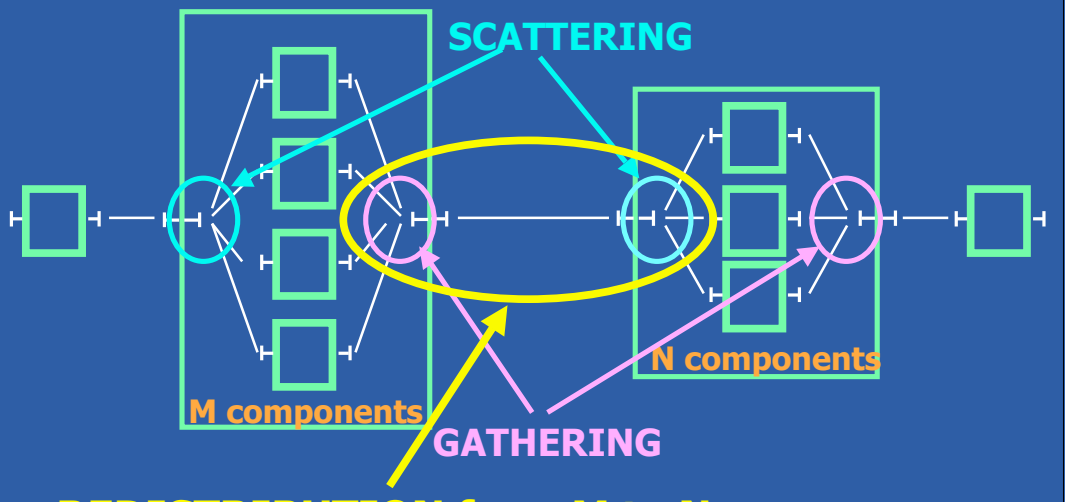
Grid on 5 clusters (DAS 2): Speed up of 100 on 150 machines

Fortran: no more than 40 proc. …

Beating Fortran MPI with Java ProActive? X/40 (14/16) = 2X/ n (100/150)

Yes, starting at 105 machines !

On Going : M x N Communications + Redistribution



Adaptive Feature: Multi-transports layer RMI, RMI-ssh, …, Ibis, HTTP XML, ...

Adaptive choice of transport layer between:

- RMI
- ssh/RMI

Also available with static configuration:

- Ibis (TCP, Myrinet, etc.)
- HTTP
- … ssh/HTTP

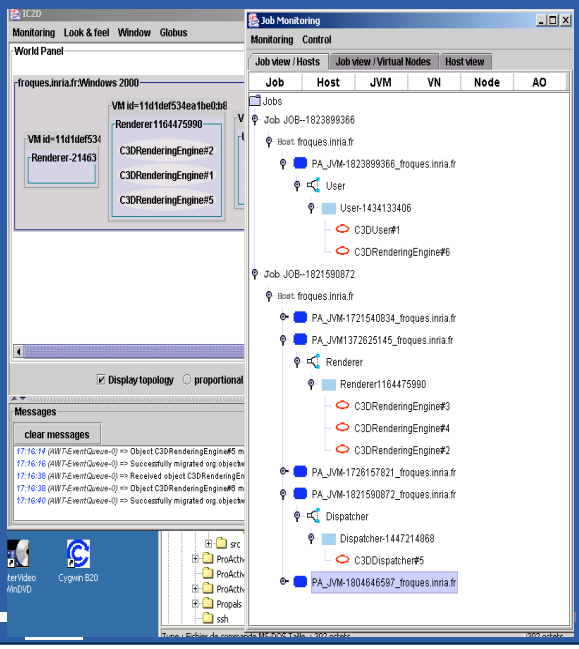
Short Term Perspective:

Fully Adaptive Choice between all transports

IC2D: Basic features cont.

Job Management:

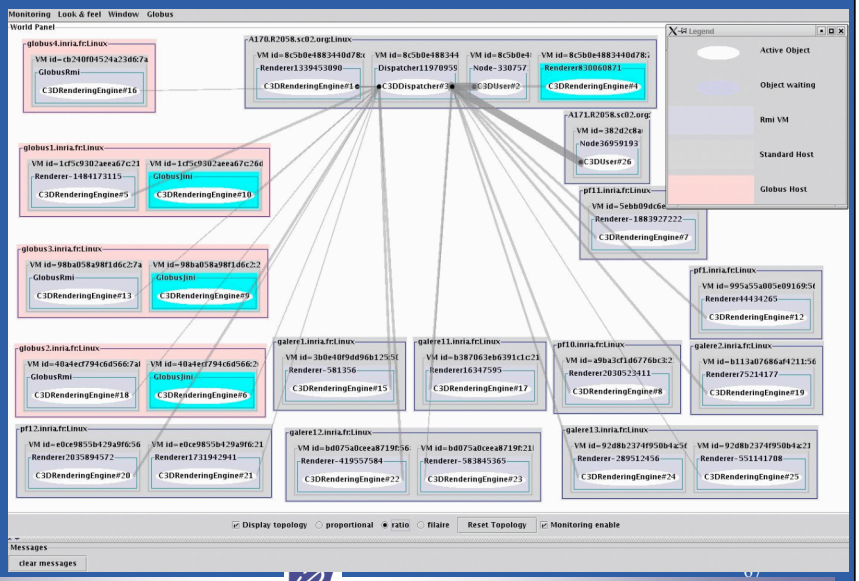
- JVM, AO per Job ID
- Textual visualisation,
- control (kill all, etc.)



Monitoring of RMI, Globus, Jini, LSF cluster Nice -- Baltimore

ProActive IC2D:

Width of links proportional to the number of com-munications



ProActive:

A Java API + Tools for Parallel, Distributed Computing

- A uniform framework: An Active Object pattern
- A formal model behind: Determinism, Insensitivity to deployment

Programming Model:

- **Remote Objects (Classes, not only Interfaces, Dynamic)**
- **Asynchronous Communications, Automatic dataflow synchro: Futures**
- **Groups, Mobility, Components, Security**

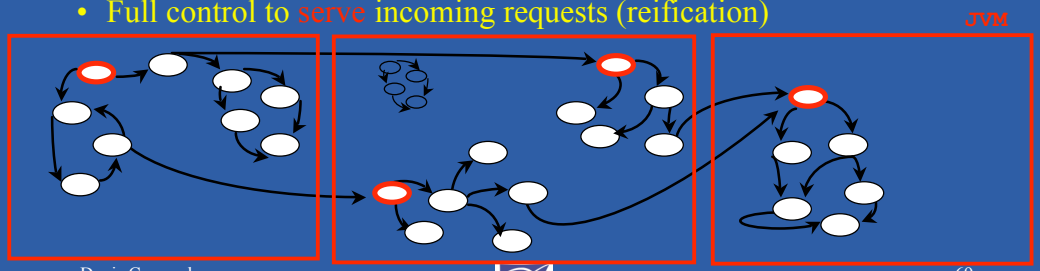
Environment:

- **XML Deployment Descriptors**
- **Interfaced with various protocols:** `rsh,ssh,LSF,Globus,Jini,RMIregistry`
- **Visualization and monitoring: IC2D**

In the www. ObjectWeb .org Consortium (Open Source middleware) since April 2002 (LGPL license)

ProActive : model

- Active objects : coarse-grained structuring entities (subsystems)
- Each active object: - possibly owns many passive objects
  - has exactly one thread.
- No shared passive objects -- Parameters are passed by deep-copy
- Asynchronous Communication between active objects
- Future objects and wait-by-necessity.
- Full control to serve incoming requests (reification)



ProActive model (2)

Java RMI (Remote Method Invocation = Object RPC = o.foo(p))

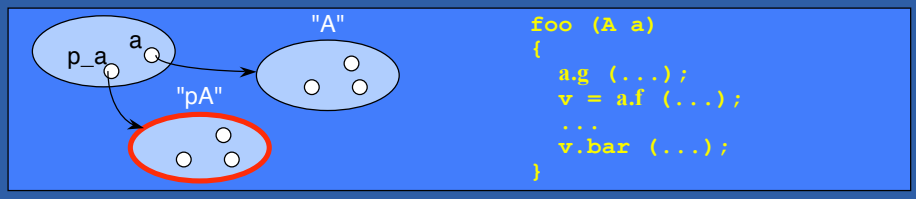
plus a few important features:

- Sequential Object: a single thread with FIFO service
- Asynchronous Method calls towards Active Objects:
  Implicit Futures as method results
- Wait-By-Necessity:
  - Automatic wait upon a strict operation on an unknown future
  - First-Class Futures:
    - Futures can be passed to other activities
    - Sending a future to another machines is not blocking

ProActive : Reuse and seamless

Two key features:

- Polymorphism between standard and active objects
  - Type compatibility for classes (and not only interfaces)
  - Needed and done for the future objects also
  - Dynamic mechanism (dynamically achieved if needed)



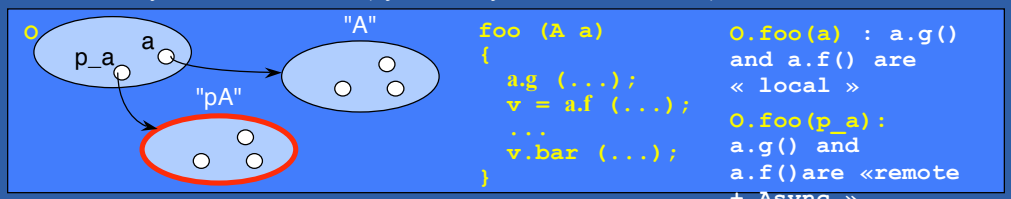
- Wait-by-necessity: inter-object synchronization
  - Systematic, implicit and transparent futures

    Ease the programming of synchronizations, and the reuse of routines

ProActive : Reuse and seamless

Two key features:

- Polymorphism between standard and active objects
  - Type compatibility for classes (and not only interfaces)
  - Needed and done for the future objects also
  - Dynamic mechanism (dynamically achieved if needed)



- Wait-by-necessity: inter-object synchronization
  - Systematic, implicit and transparent futures (“value to come”)

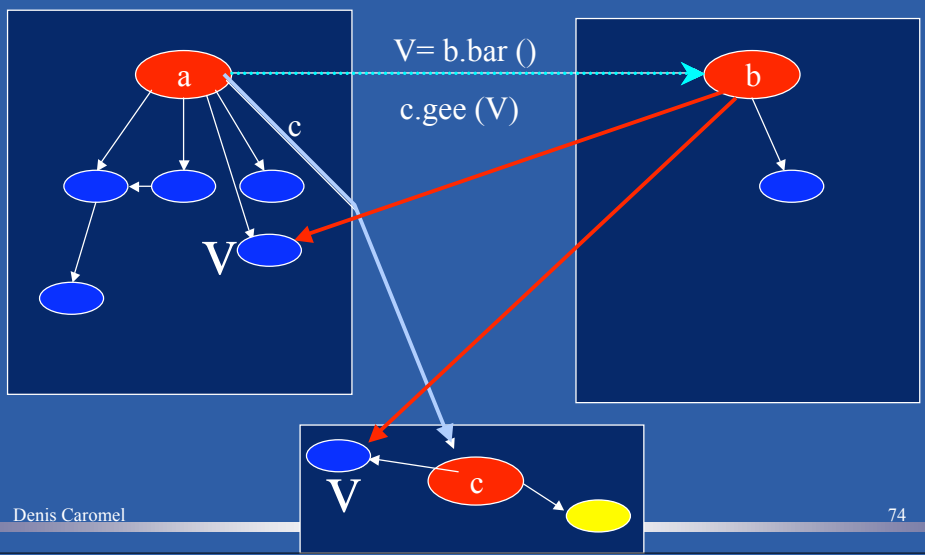
    Ease the programming of synchronizations, and the reuse of routines

First-Class Futures

Update

Wait-By-Necessity: First Class Futures

Futures are Global Single-Assignment Variables



Future update strategies

No partial replies and requests:

- No passing of futures between activities, more deadlocks

Eager strategies: as soon as a future is computed

- Forward-based:
  - Each activity is responsible for updating the values of futures it has forwarded
- Message-based:
  - Each forwarding of future generates a message sent to the computing activity
  - The computing activity is responsible for sending the value to all

Mixed strategy:

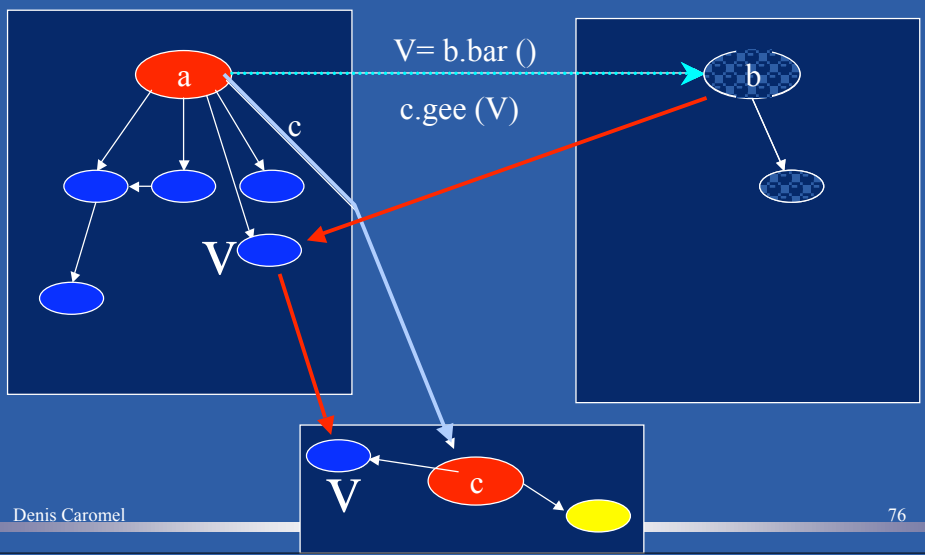
- Futures update any time between future computation and WbN

Lazy strategy:

- On demand, only when the value of the future is needed (WbN on it)

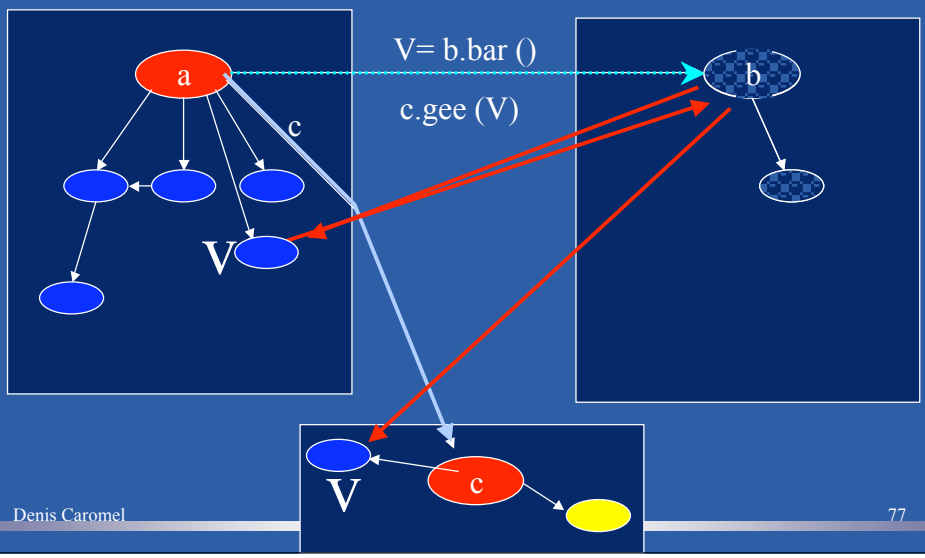
Wait-By-Necessity: Eager Forward Based

AO forwarding a future: will have to forward its value



Wait-By-Necessity: Eager Message Based

AO forwarding a future: send a message



Wait-By-Necessity: Lazy Strategy

An Active Object requests a Future Value when needed