



Architect of an Open World™

Extrapolation et Estimation de performance

dans le cadre des réponses aux appels d'offre

LIBERATE IT

1969 – ...

Xavier VIGOUROUX

Marié

3 enfants

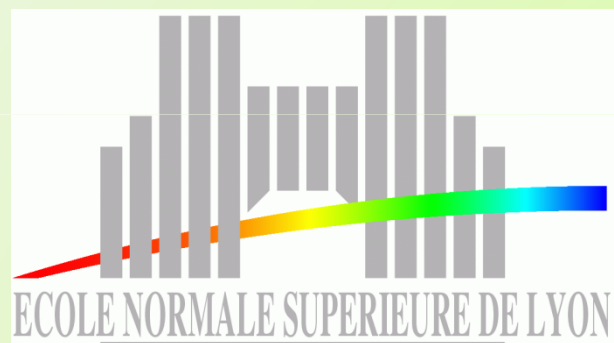
1987 – 1989

Université Claude Bernard



Lyon 1

1989 – 1996



Jean-Marc Vincent





Janvier 1996

Analyse distribuée de traces d'exécution de
programmes parallèles.

1997 - 2000



i n v e n t

2000 - 2006



2006 - ...

Responsable de l'équipe HPC benchmark



Mon Travail

Le contexte industriel

Le but de Bull est de vendre des machines.

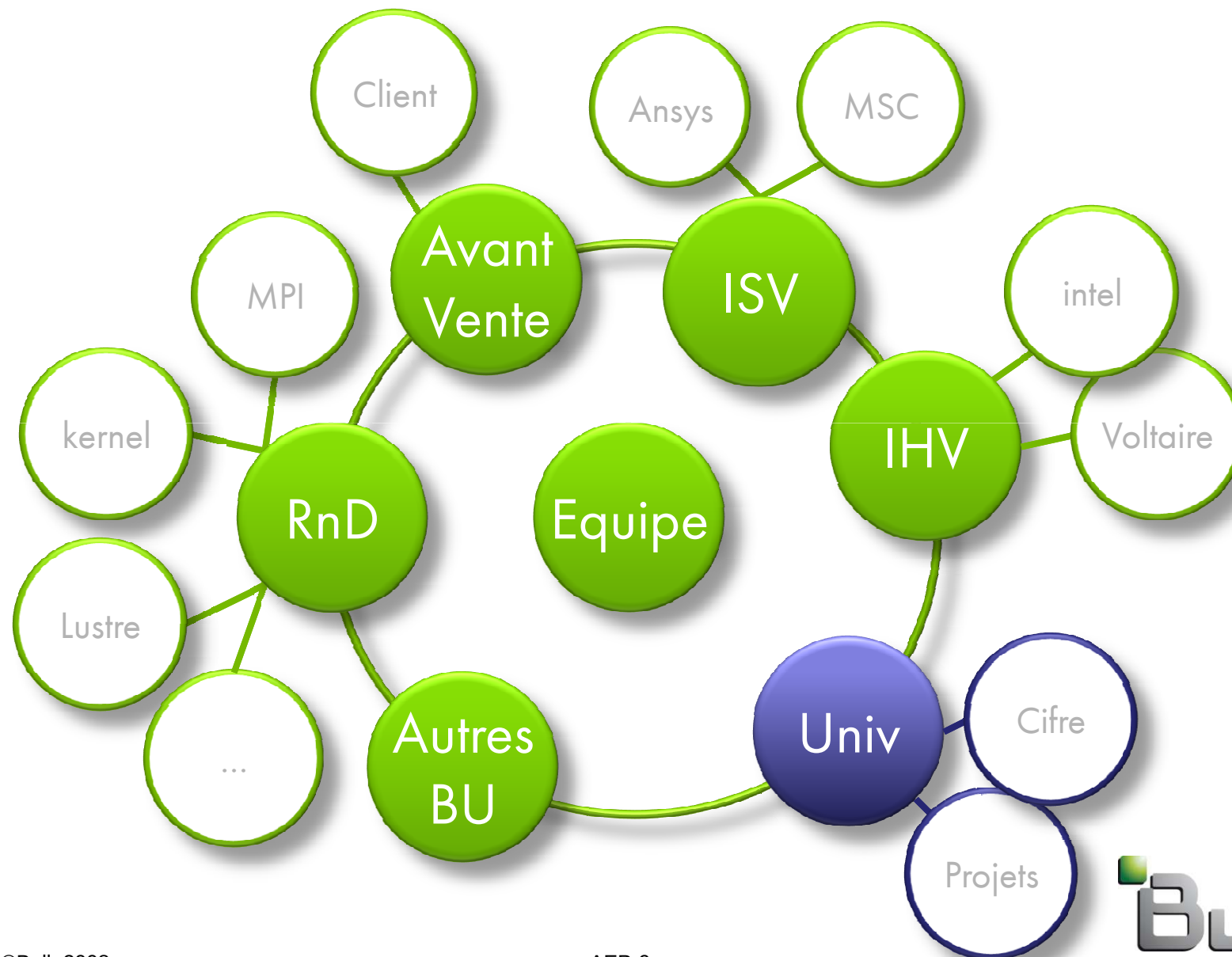
A ressource contrainte, il faut satisfaire le
court-terme et le long-terme

Je suis le court-terme

Vous êtes le long-terme



Autour de l'équipe



Mes activités

Avant-vente

Portage
Benchmark

Démo

Visite

Accès

Delivery

Staging

Acceptance

Fast Start

Après-vente

Training

Optimisation

Support L3



L'appel d'offre sur performance

- Principe

- Proposer la machine maximisant les performances des codes client compte tenu des contraintes

- Problématique

- Temps de réponse relativement court (2 jours à 6 mois)
- Disponibilité du matériel : matériel non encore introduit, taille de la machine trop importante.
- Roadmap logiciel
- Engagement aujourd'hui pour le jour de la livraison (délais entre 3 mois et 2 ans).



Comment nous engager ?

- Dans un monde idéal
 - Le temps n'est pas limité
 - Une étude exhaustive des profiling des codes est possible (inter/intra nœud)
 - Une campagne de mesure est possible
 - Nous avons un modèle d'extrapolation pour les machines futures
- Dans notre monde
 - Le temps est extrêmement limité
 - Les machines ne sont pas toujours disponibles
 - Les modèles sont rares pour les processeurs futur....

Les risques

	Engagements		
	Sous-estimés	Bonne estimation	Sur-Estimés
On Gagne	Faut pas rêver!	Parfait	500k€ par critère
On Perd	On me tombe dessus	C'est la vie	Tant pis pour la concurrence

Déroulement d'un appel d'offre

1. Réception de l'A.O. et du Cahier des Clauses Techniques Particulières

- Définition des benchmarks et des conditions de passage

2. Définition d'une architecture

- Type de processeur, nombre de cœurs
- Architecture du réseau d'interconnect
- Architecture du système d'entrée/sortie

3. Passage des benchmarks

4. Extrapolation/Estimation

5. Rédaction du rapport pour la réponse à l'A.O.



Exemple de demande...

- Une machine d'une capacité de 10 Petaflops
- Dans un espace d'un demi terrain de football
- D'une consommation de 1 MWatt
- Engagement de performance sur 20 benchmarks
- A livrer pour 2010
- Réponse à fournir pour dans 15 jours



Nos moyens pour répondre

- Nos moyens

- Une machine d'une capacité d'un petaflops
- Des processeurs de la génération précédente
- Des machines dont l'architecture sera dépassée
- Un interconnect actuel
- Des outils: vi (ou emacs), sh, csh, bash, ksh, ..., -g -pg, gprof, ...

- Ce que nous savons

- L'architecture imaginée
- La promesse des fréquences futures des processeurs



Mes problèmes ... avec le CPU

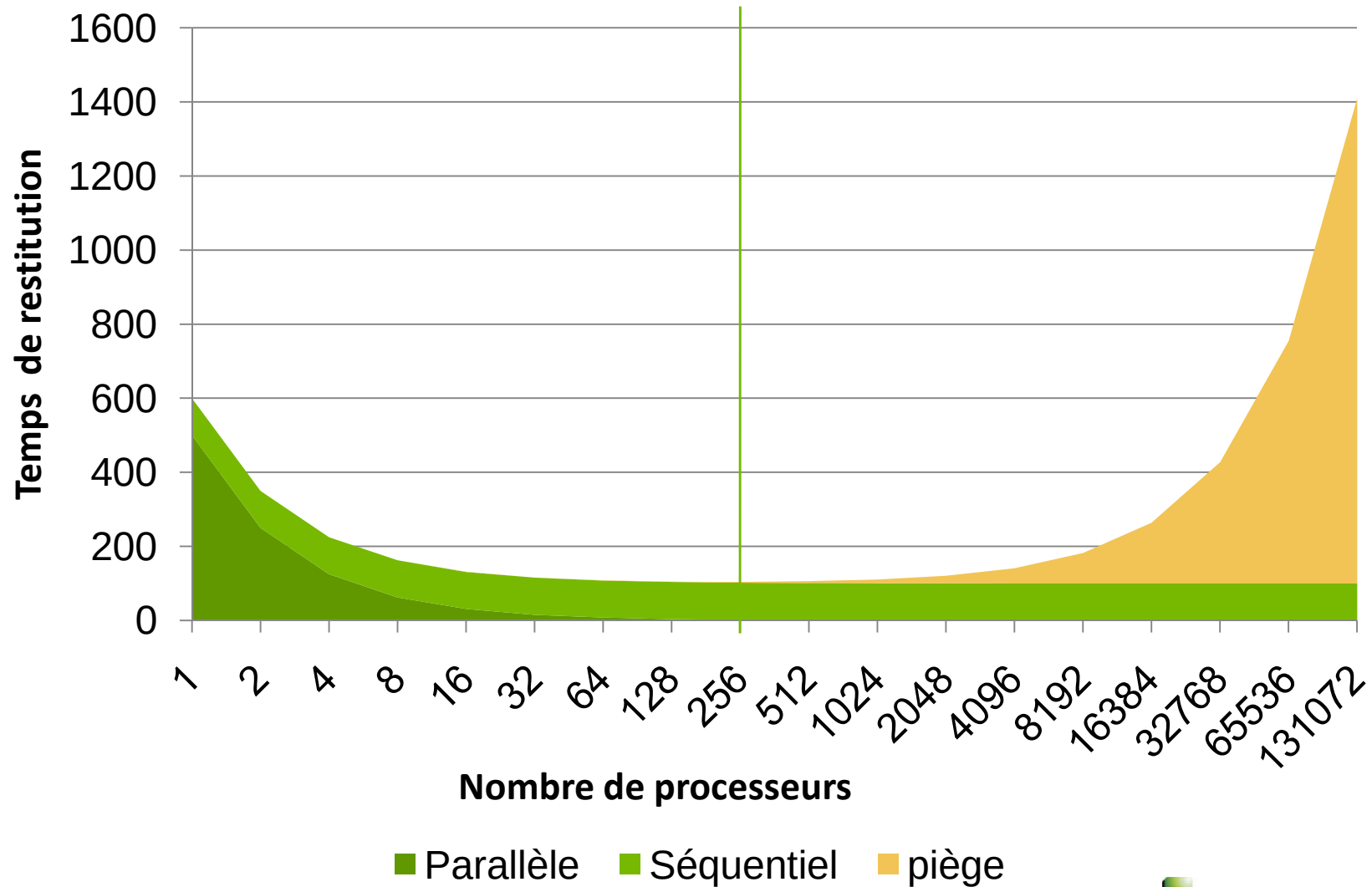
- Plus la processeur est intelligent, moins je sais ce qu'il fait.
 - Itanium vs Xeon
 - Préfetching ?

Mes problèmes ... avec la taille

l'effet cache



Et encore un autre problème....de taille



Aujourd'hui, il faut être honnête !

Nous n'extrapolons pas les performances...
Nous les estimons.

Nous nous basons sur :
nos expériences, les quelques mesures possibles

Mes questions existentielles

- « Comment se comportera ce code sur plus de cœurs ? »
- « Comment se comportera ce code sur avec ce processeur ? »
- « En quoi la bande passante mémoire est limitante pour ce code? »
- « Quel sera le temps d'exécution en mode dépeuplé ? »
- « La latence réseau est-elle importante pour ce code ? »
- « Quelle taille mémoire est nécessaire ? »

De quels modèle ai-je besoin ?

Celui qui répond à la question:

Le temps de restitution de ce cas test
sur la machine X sera de Y secondes.

Je ne rêve pas

Je suis d'accord pour ne changer qu'un paramètre à la fois:

- Bande passante mémoire (saturation ?)
- Fréquence CPU (préfetching ? CPU bounded ?)
- Latence réseau (petits paquets ?)
- Bande passante réseau (gros paquets ?)



Futur et Rupture

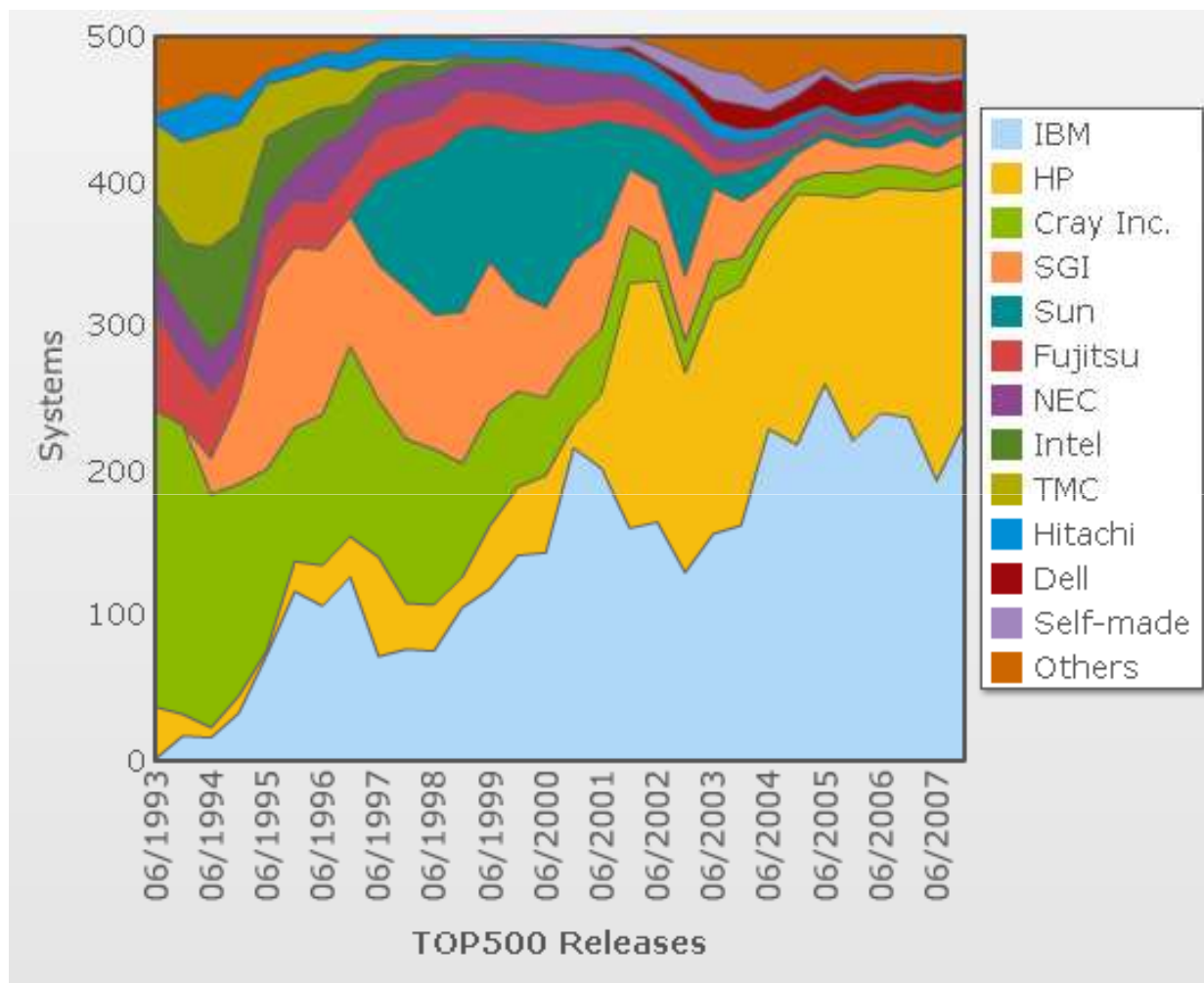
De quoi est fait le HPC

- Le HPC ne se suffit pas à lui-même.
- Il ne peut que reposer sur des composants standards
 - Ces composants sont sur un autre marché
 - Ces composants vont être sur un autre marché (IB et le stockage)

La preuve ?



Parts de marché dans le top500



A parte - Performance Development



Nous nous dirigeons vers du « tout standard »

Réseau

Nœuds

CPU

IO

- Une seule topologie: Fat tree
- Fin des Opérations collectives (quadrics)
- Montée du Gigabit Ethernet.
- IP over IB

Nœuds et CPU standard, standard, standard

HP: fin du PA-RISC pour intel

IBM: power et AMD (xseries)

Sun: SPARC et AMD et intel

Cray: collaboration avec intel !

SGI: intel

NEC: earth simulator



Les ruptures viendront des grandes industries

Si on regarde intel....

- On voit déjà
 - Multi-core
 - Intégration du contrôleur mémoire dans le CPU
 - Hétérogénéité extra-CPU (GPGPU, FPGA)
- On voit arriver
 - Stacking
 - Multi-cœur hétérogène
 - Vectorisation





Ma liste au père Noël

Je veux

- Des modèles simples
- Des modèles utilisables
- Des modèles prenant en compte « les ruptures »
- D'abord en ordre dispersé...puis s'intégrant les uns aux autres (à la code météo)

Je ne veux pas

- Un modèle demandant 40 paramètres que je ne sais pas récupérer
- intrusif





Architect of an Open World™

Remerciement à Dick Hardt – CEO SXIP

