



# KAUST is a research institution with four sustainability "pillars":



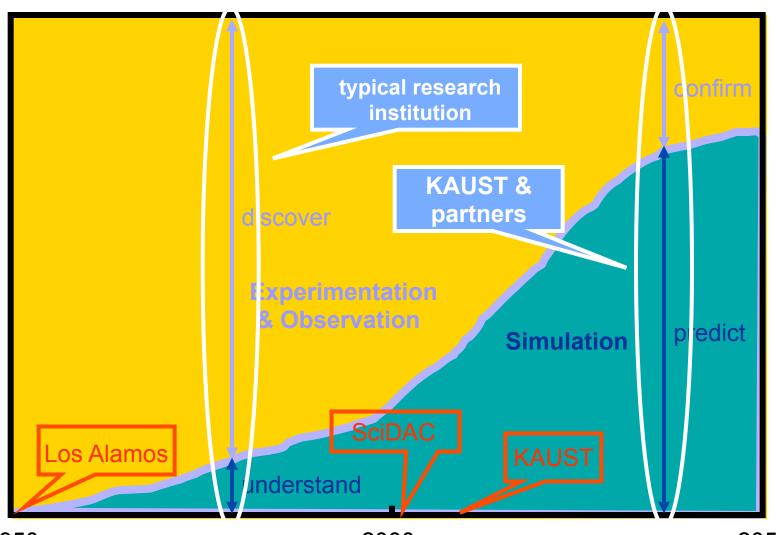
... and computational science as an "enabler"

# ... with four paradigms to employ



(many institutions are still firing on just two of the four cylinders)

# Advance of the third paradigm



1950 2000 2050



# KAUST's overall strategy: leapfrog

#### KAUST faces two means to achieve its mission of

- Advancing science and technology
- Catalyzing economic diversification
- Connecting Saudi Arabia to best practices in research and academic culture

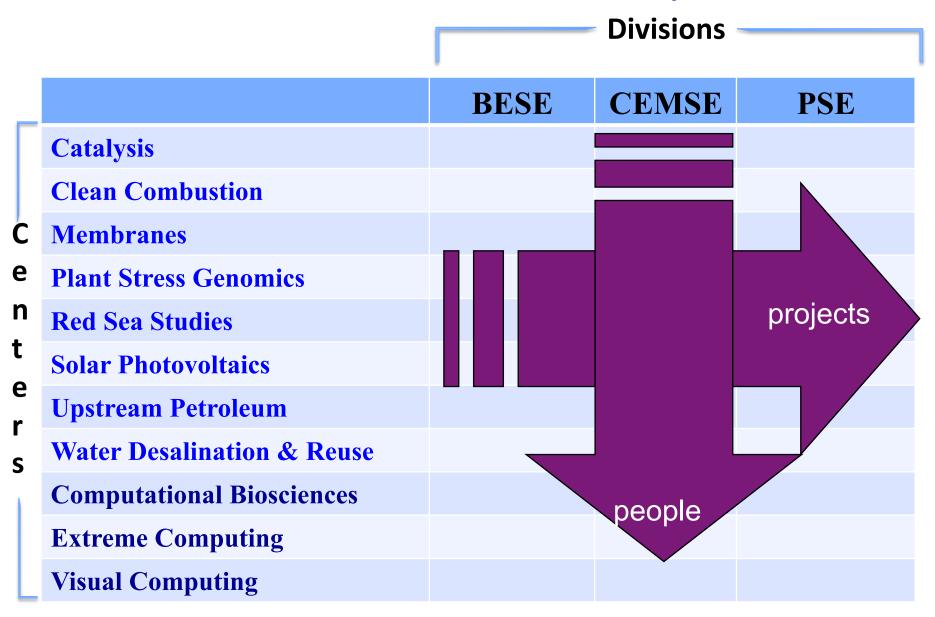
### 1. Patiently follow the leaders

- Do what they do, long and well, or
- 2. "Reinvent" the research university
  - Emphasize what others may find less natural, being "trapped" in their success

# **KAUST's strategy #1: Multidisciplinarity**

- KAUST is structured to capitalize on neglected ground between traditional disciplines
- Science and engineering are administered together as equals
- Mission-driven research centers draw expertise required from any academic unit, by adding independent resources
- Disciplinary territoriality is reduced
  - no student credit hour-based budgeting
  - faculty lines assigned to serve missions

## KAUST's matrix structure, 2015



# KAUST's strategy #2: Big Sisters degree program founding partners

Partner	Program
Stanford	Computer Sci
Stanford	Applied Math & Computational Sci
Berkeley	Mechanical Eng
Michigan	Electrical Eng
Texas	Earth Sci & Eng
Illinois	Environmental Sci & Eng
Woods Hole*	Marine Sci & Eng
Imperial College	Materials Sci & Eng
Cambridge	Biological Sci & Eng

<sup>\*</sup> small partner

# KAUST's strategy #2: Big Sisters research program founding partners



# **KAUST's strategy #3: Industry and Entrepreneurship**



































































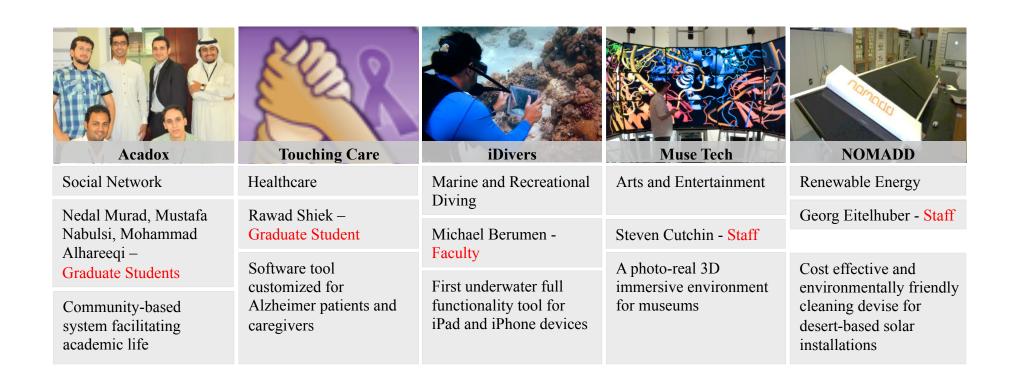




# KAUST's strategy #3: Industry and Entrepreneurship

- KAUST aims that regional industrial demand will drive much on-campus research
- Faculty, post-docs, students are encouraged and enabled to be entrepreneurs with
  - short courses
  - industrial partners in residence
  - seed funding
  - expert staffing for IP, fund raising, etc.

# KAUST's strategy #3: Industry and Entrepreneurship Student and staff start-ups



4 of first 5 start-ups are in domain of information technology

# **KAUST's strategy #4: Facilities**

#### Seven state-of-the-art core labs

- Supercomputing
- Scientific visualization
- Nanofabrication
- DNA sequencing
- Imaging and characterization
- Marine science
- Machine fabrication



**KAUST's strategy #4: Facilities** 







# Shaheen II Specs

- 36 cabinets of Cray XC40 with Intel Haswell 2.3 Ghz with 16 cores

- 128 GB of RAM per node

- Number of nodes: 6192

- Number of cores: 198144

- Peak Performance: 7.3 PFlops/s

- LINPACK: 5.6 PFlops/s

- 2.8 MW at peak

- 17.4 PB of Parallel File System

- I/O throughput: over 500 GB/s

- Burst Buffer capacity: 1.5 TB

- Burst Buffer throughput: over 1.2 TB/s

#### Newly ranked: #7 on HPL

#7 on HPCG

## ~2 GF/s/W



#### Society for Industrial and Applied Mathematics









SIAM NEWS >

#### CSE 2009: The World's First CSE University

June 15, 2009

The King Abdullah University of Science and Technology, scheduled to welcome its first class of students in September, sponsored a reception in Miami on March 2, the first day of the SIAM Conference on Computational Science and Engineering. David Keyes and Omar Ghattas, involved in different ways in the new venture, hosted the reception and made informal presentations to the assembled crowd.

Most readers will know something of KAUST, which for the record is a graduate-only (master's and doctoral) university being constructed in Saudi Arabia, on the eastern edge of the Red Sea, not far from Jeddah. Keyes, the inaugural chair of KAUST's Mathematical and Computer Sciences and Engineering Division, offered examples of research areas of particular interest to Saudi Arabia and the region that will be emphasized; among them are geophysics, seismology, reservoir modeling, CO2 sequestration, photovoltaics, stress-tolerant agriculture, desalination, catalysis, and materials, along with the applied mathematics and computer science required to support them.

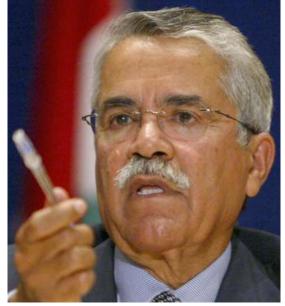
Sizeable recruitment ads for KAUST have appeared in many recent issues of SIAM News often side by side with ads placed by partners of the new university, such as the

# How we got here

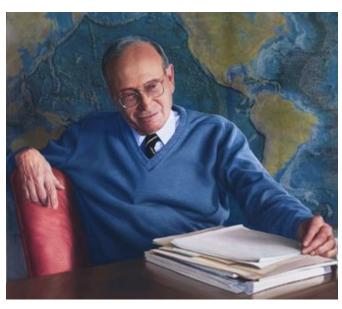
The VISION



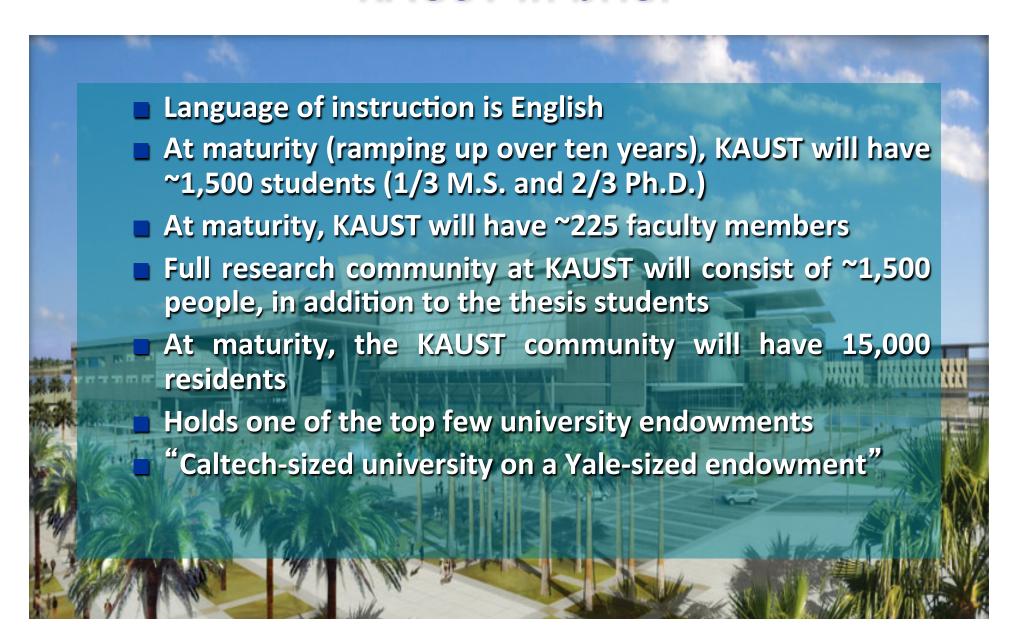
The EXECUTION



The DESIGN



## **KAUST** in brief



## **KAUST** in brief

Graduate university governed by independent, self-perpetuating Trustees, operated off of an independently held endowment, with Western-style charter

Open to men and women from all over the world

Organized as a matrix of academic divisions and research centers

Globally connected at creation through KAUST-funded partnerships

Engaged in collaborative research with the private sector to advance economic development

Occupying a state-of-the-art, LEED-certified, environmentally responsible campus

# 2015 Leiden size-independent rankings

### http://www.leidenranking.com ( ) THOMSON REUTERS

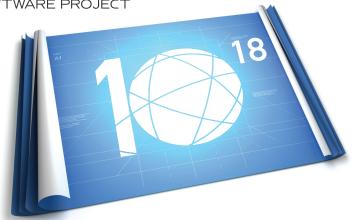


Rank	University	Country	Rank	University	Country	Rank	University	Country
1	MIT	700	26	Univ Massachusetts Med Sch		51	Vanderbilt Univ	
2	Harvard Univ	200	27	Univ Washington - Seattle		52	Leiden Univ	
3	Stanford Univ	200	28	Cornell Univ		53	Univ Southern Calif	100
4	Univ Calif - Berkeley	200	29	Washington Univ - St Louis		54	<u>Univ Dundee</u>	開題
5	Princeton Univ	200	30	NYU		55	Univ Aberdeen	
6	Caltech	200	31	<u>Duke Univ</u>		56	Univ Wisconsin - Madison	200
7	Univ Calif - Santa Barbara	200	32	Univ Coll London		57	Emory Univ	
8	Univ Calif - San Francisco	200	33	Imperial Coll London		58	Univ Maryland - College Park	
9	Rice Univ	203	34	Univ Exeter		59	Dartmouth Coll	
10	Weizmann Inst Sci	0	35	King's Coll London		60	Utrecht Univ	
11	London Sch Hyg & Trop Med		36	Johns Hopkins Univ		61	Univ Geneva	•
12	Univ Calif - Santa Cruz	200	37	Icahn Sch Med - Mt Sinai		62	Univ Massachusetts Amherst	
13	Yale Univ	200	38	Boston Coll		63	Univ Edinburgh	
14	Univ Texas - Southwestern Med Ctr	203	39	Univ N Carolina - Chapel Hill		64	VU Univ Amsterdam	
15	Ecole Polytech Fed Lausanne	•	40	Univ Basel	•	65	Carnegie Mellon Univ	
16	Univ Calif - San Diego	200	41	<u>Univ Lausanne</u>	•	66	École Polytech	
17	Univ Oxford		42	Boston Univ		67	Univ Illinois - Urbana-Champaign	
18	Univ Chicago	200		Univ Bristol		68	<u>Univ Amsterdam</u>	_
19	Columbia Univ	200	44	<u>Univ Calif - Riverside</u>		69	<u>Univ York</u>	
20	Univ Calif - Los Angeles	200	45	Univ St Andrews		70	Tufts Univ	
21	Northwestern Univ	200	46	Univ Calif - Irvine		71	Katholieke Univ Leuven	•
22	Univ Colorado - Boulder	200	47	Univ Texas - Austin		72	Paris Descartes Univ	
23	Univ Cambridge		48	Univ Michigan		73	Yeshiva Univ	
24	<u>Univ Penn</u>	203		Georgia Inst Technol	<b>200</b>	74	Univ Calif - Davis	
25	ETH Zurich	•		Univ Texas Hlth Sci Ctr - Houston		75	King Abdullah Univ Sci & Technol	Eng.



# Roadmap for ECRC research





Jack Dongarra
Pete Beckman
Terry Moore
Patrick Aerts
Giovanni Aloisio
Jean-Claude Andre
David Barkai
Jean-Yves Berthou
Taisuke Boku
Bertrand Braunschweig
Franck Cappello
Barbara Chapman
Xuebin Chi

Alok Choudhary Sudip Dosanjh Thom Dunning Sandro Fiore Al Geist Bill Gropp Mark Hereld Michael Heroux Adolfy Hoisie Koh Hotta Yutaka Ishikawa Fred Johnson



Jesus Labarta Alain Lichnewsky Thomas Lippert Bob Lucas Barney Maccabe Satoshi Matsuoka Paul Messina Peter Michielse Bernd Mohr Matthias Mueller Wolfgang Nagel Hiroshi Nakashima Michael E. Papka Dan Reed Mitsuhisa Sato Ed Seidel John Shalf David Skinner Marc Snir Thomas Sterling Rick Stevens

Fred Streitz

Bob Sugar Shinji Sumimoto William Tang John Taylor Rajeev Thakur Anne Trefethen Mateo Valero Aad van der Steen Jeffrey Vetter Peg Williams Roter Wisniewski Kathy Yelick

SPONSORS





























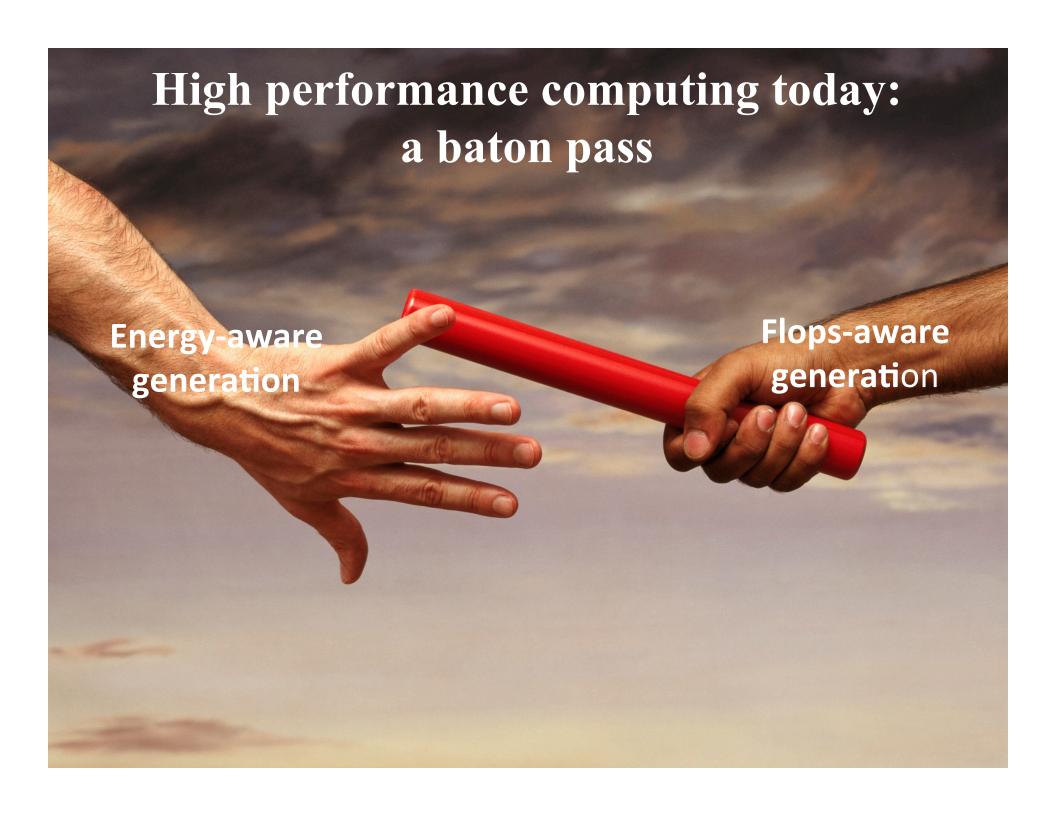






The International Exascale Software Roadmap, J. Dongarra, P. Beckman, et al., International Journal of High Performance Computer Applications **25**(1), 2011, ISSN

1094-3420.



### **ECRC** current students



Ali Lebanon



**Amani** Saudi



Chengbin China



Dalal Jordan



Gustavo **Mexico** 



Huda Jordan



Lulu China



**Mohammed** Saudi



Mustafa **Syria** 



Noha Saudi



**Tareq** Jordan



Wajih Lebanon

# ECRC student or post-doc alumni



**Ahmad** 



Aron



Enas **UTenn USArmy ERDC OSU** 



Jennifer **UManchester** 



Kai SINTEF



**Xiangliang KAUST** 



Yuxin (tbd)

### Other faces of the ECRC



George Turkiyyah Research Scientist



Stefano Zampini Research Scientist



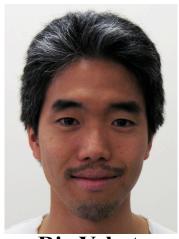
Hatem Ltaief Research Scientist



Alex Litvinenko Research Scientist



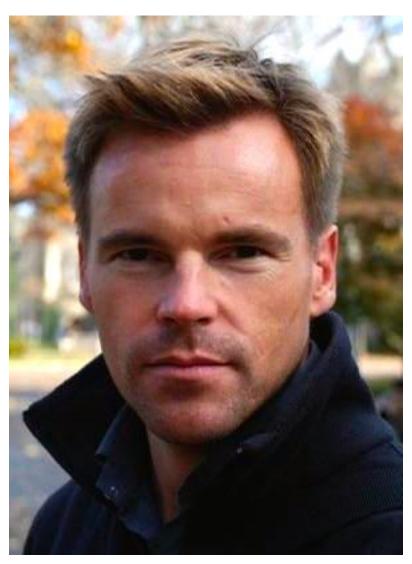
Valerio Pascucci Visiting Professor



Rio Yokota Visiting Professor

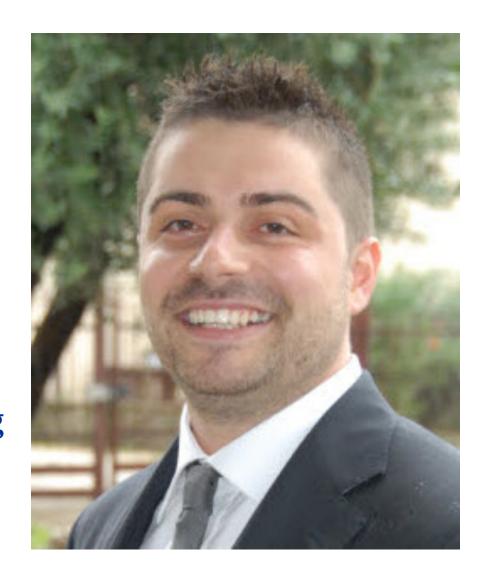
# ECRC's newest faculty appointees

- Daniel Peter
- PhD, ETH, 2008
- Numerical methods for seismic wave propagation
- Applications of high performance computing
- Architect of SPECFEM3D\_GLOBE



# ECRC's newest faculty appointees

- Matteo Parsani
- PhD, Frije Universiteit
   Brussel, 2010
- Numerical methods for compressible and reacting flows
- Applications of high performance computing
- Co-designer of NASA's
   FUN3D design code



# KAUST's PETSc developers

#### **PETSc Team**

- Home
- Download
- Features
- Documentation
- Applications/Publications
- Miscellaneous
  - PETSc Team
  - o Mailing Lists
  - Acknowledgements
  - Funding
  - Threads and PETSc
- External Software
- Developers Site

Please do not email individual PETSc authors. E-mail to <a href="mailto:petsc-maint@mcs.anl.gov">petsc-maint@mcs.anl.gov</a> is automatically distributed to the authors, so our response time using this address will be fastest.

#### **Current Active Developers (August 2014)**



**Shrirang Abhyankar** 



Mark Adams



Satish Balay



Jed Brown



Lisandro Dalcin



**Tobin Isaac** 



**Dmitry Karpeev** 



Matthew Knepley



**Lois Curfman** 



**Karl Rupp** 



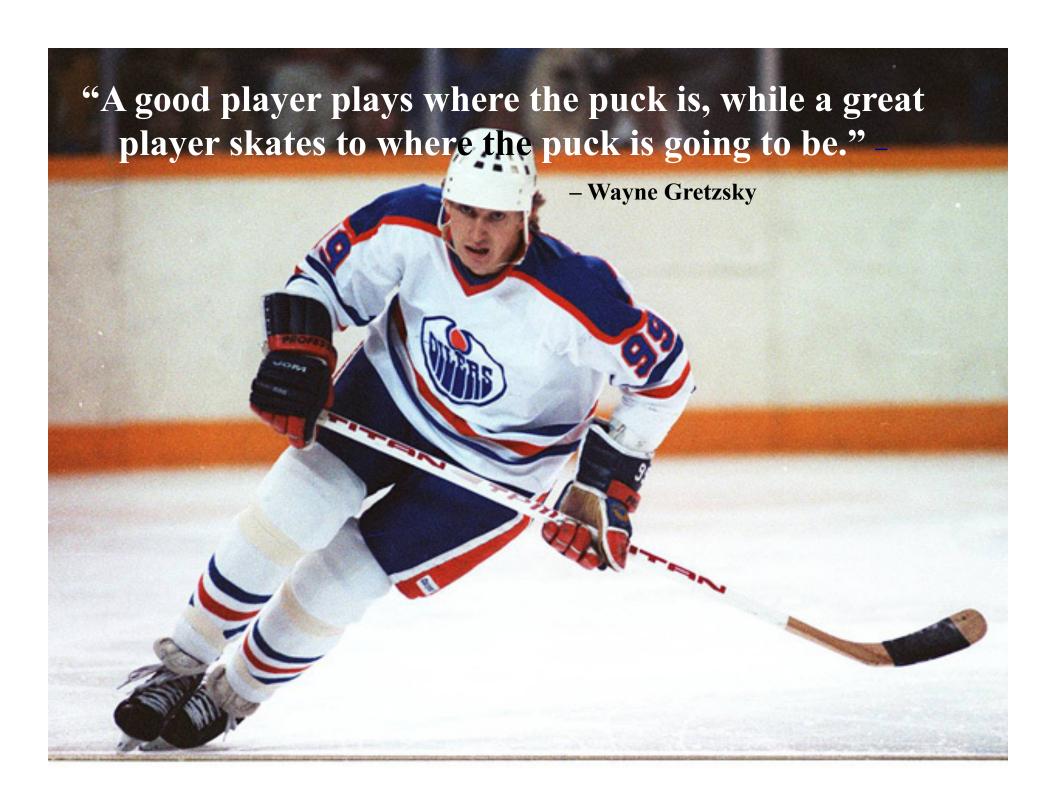
Jason Sarich



**Barry Smith** 



Hong Zhang



# **ECRC** aspiration

To paraphrase Gretzsky:

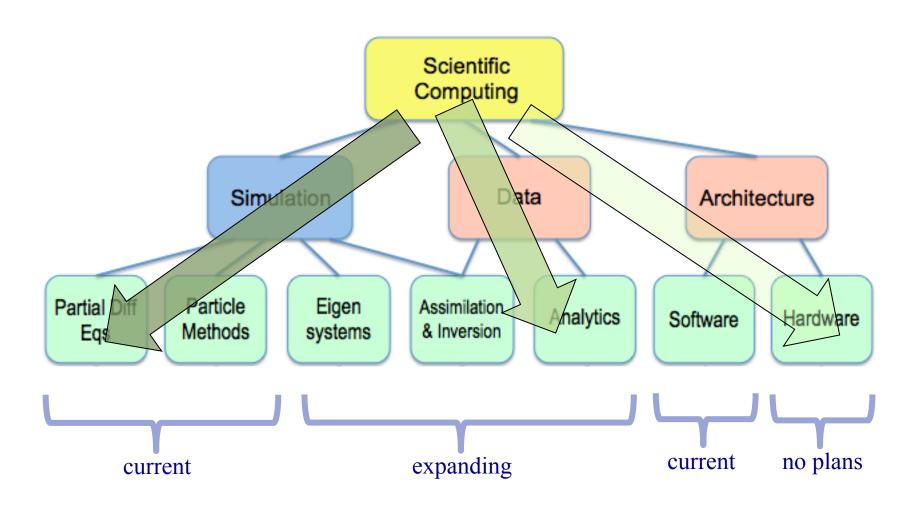
"Algorithms for where architectures are going to be"

Such algorithms may or may not be the best today; however, hardware trends can be extrapolated to their sweet spots.

#### ECRC's twin deliverables

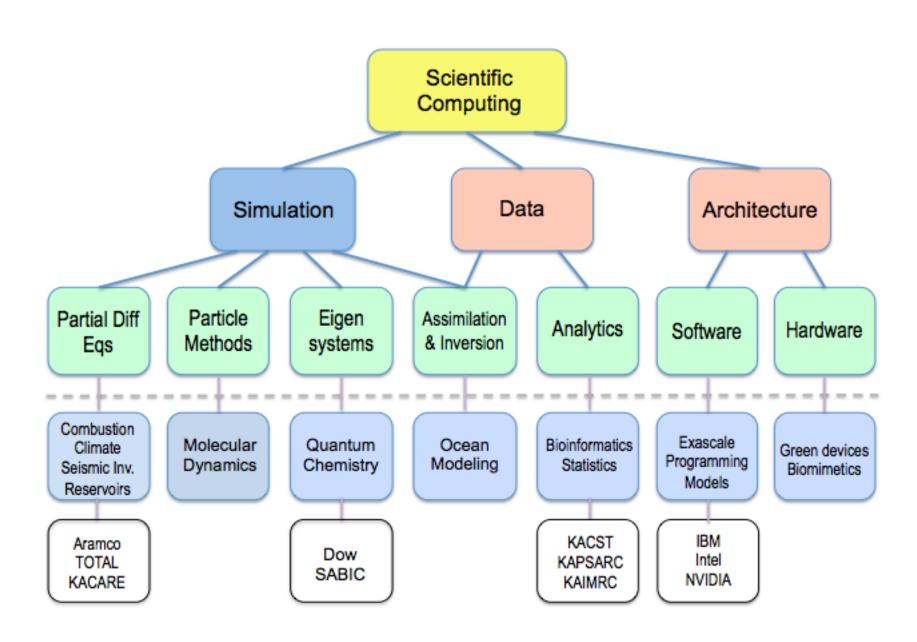
- Develop and demonstrate algorithms for scientific simulation and data analytics at emerging architectural scales
  - most of today's scientific libraries are designed for a different environment (BSP) and need to be evolved
- Raise the impact and increase the reach of scientific simulation and data analytics at KAUST
  - pursuit of performance requires specialization complementary to expertise of computational scientists and engineers

# Vision for extreme computing at KAUST



Topical foci of the ECRC relative to the spectrum of scientific computing at KAUST today

# Vision for extreme computing at KAUST



# Required software enabling technologies

### **Model-related**

- Geometric modelers
- Meshers
- Discretizers
- Partitioners
- Solvers / integrators
- Adaptivity systems
- Random no. generators ■
- Subgridscale physics
- Uncertainty quantification
- Dynamic load balancing
- Graphs and combinatorial algs.
- Compression

### **Development-related**

- Configuration systems
- Source-to-source translators
- Compilers
- Simulators
- Messaging systems
- Debuggers
  - **Profilers**

High-end computers come with little of this stuff. Most has to be contributed by the user community

### **Production-related**

- Dynamic resource management
- Dynamic performance optimization
- Authenticators
- I/O systems
- Visualization systems
- Workflow controllers
- Frameworks
- Data miners
- Fault monitoring, reporting, and recovery

### Relation to KAUST's economic mission

- Enable advances through simulation for the KSA's largest industry today:
  - seismic imaging for petroleum discovery
  - reservoir simulation for petroleum production
  - computational chemistry for downstream petroleum processing
- Attract an important industry the KSA does not possess today:
  - information technology







### Collaborations within KAUST

- Obvious center collaborations:
  - Combustion, Catalysis, Red Sea
- Possible center collaborations:
  - Water Desalination, Membranes, Computational Biosciences, Solar & Renewables
- Core laboratories:
  - Supercomputing Laboratory
  - Visualization Laboratory

### Collaborations within the KSA

- Agencies/ministries:
  - KACARE, KACST, KAIMRC, KAPSARC
- Universities:
  - KFUPM, UQU
- Companies:
  - Aramco, SABIC

### Collaborations outside the KSA

- Computational science and computer science programs at national laboratories of the U.S. DOE and European PRACE
- IT vendors:
  - Cray, IBM, Intel, NVIDIA, and others
- IT customers:
  - Total and others
- EU Horizon2020 institutions
- Other international universities

# Algorithmic agenda

- New formulations with
  - greater arithmetic intensity (flops per byte moved into and out of registers and upper cache)
    - including assured accuracy with (adaptively) less floatingpoint precision
  - reduced synchronization and communication
    - less frequent and/or less global
  - greater SIMD-style thread concurrency for accelerators
  - algorithmic resilience to various types of faults
- Quantification of trades between limited resources
- Plus all of the exciting analytical agendas that exascale is meant to exploit
  - "post-forward" problems: optimization, data assimilation, parameter inversion, uncertainty quantification, etc.

# Optimal hierarchical algorithms

- At large scale, one must start with algorithms with optimal asymptotic scaling,  $O(N \log^p N)$
- Some optimal hierarchical algorithms
  - Fast Fourier Transform (1960's)
  - **■** Multigrid (1970's)
  - Fast Multipole (1980's)
  - Sparse Grids (1990's)
  - H matrices (2000's)

"With great computational power comes great algorithmic responsibility." – Longfei Gao

# **Examples being developed at the Extreme Computing Research Center at KAUST**

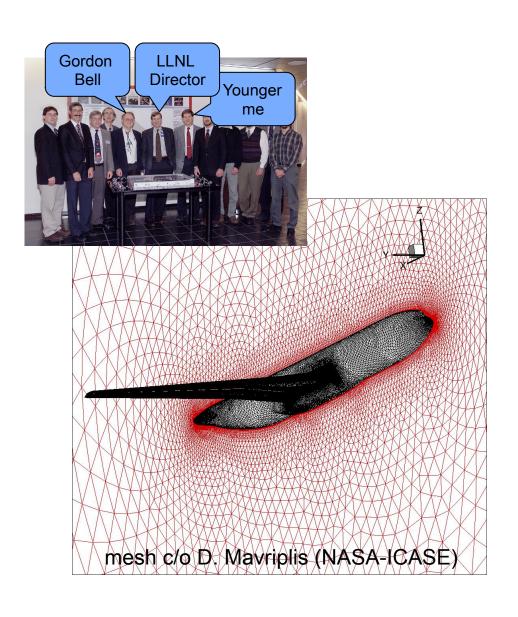
- ACR, a new spin on 45-year-old cyclic reduction that recursively uses  $\mathcal{H}$  matrices on Schur complements to reduce  $O(N^2)$  complexity to  $O(N \log^2 N)$
- FMM, a 30-year-old O(N) solver for potential problems with good asymptotics but a bad constant when used at high precision, used in low precision as a FEM preconditioner
- QDWH-SVD, a 2-year-old SVD algorithm that performs more flops but generates essentially arbitrary amounts of dynamically schedulable concurrency, and beats state-of-the-art on GPUs
- MWD, a multicore wavefront diamond-tiling stencil evaluation library that reduces memory bandwidth pressure on multicore processors
- BDDC, a preconditioner well suited for high-contrast elliptic problems that trades lots of local flops for low iteration count, now in PETSc
- MSPIN, a new nonlinear preconditioner that replaces most of the global synchronizations of Newton iteration with local problems

### Three thumbnails of ECRC research

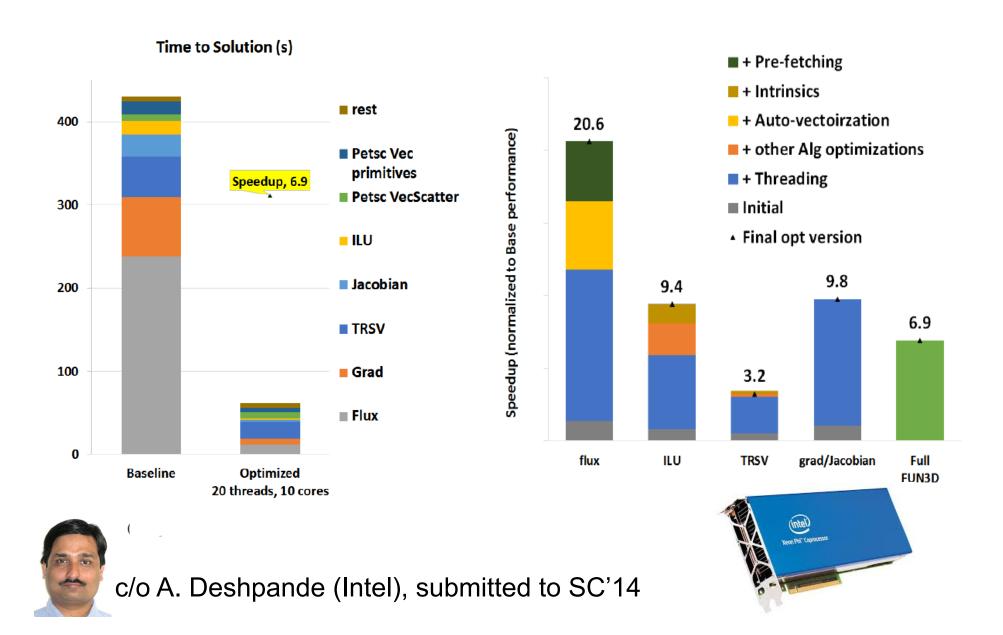
- Collaboration with Intel on aerodynamics modeling
- Collaboration with *Observatoire de Paris* on adaptive optics for world's largest telescope
- Internal algorithmic research on "algebraic fast multipole"

# Computational aerodynamics

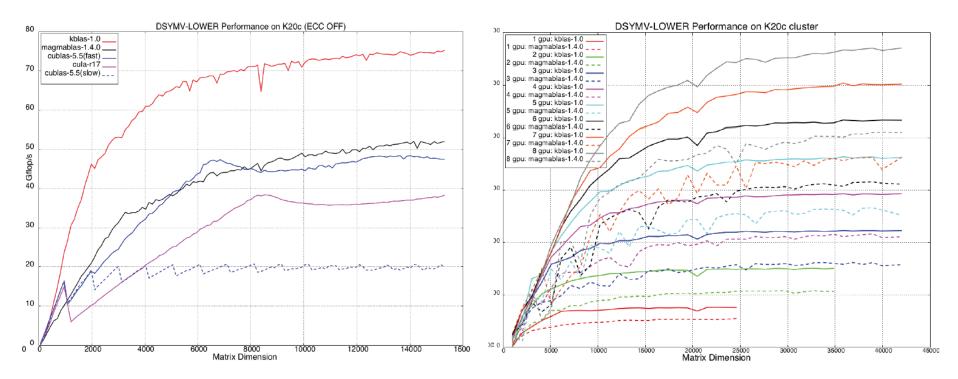
- Our PETSc-FUN3D won an ACM Gordon Bell Prize in 1999 for distributed memory scaling of a fully implicit unstructured grid NASA external aircraft flow
- With Intel-Bangalore, we are porting this implicit unstructured grid code to Intel many-core chips for strong shared-memory scaling within a single compute node



# PETSc-FUN3D on Intel "Ivy Town"



### KAUST's software in NVIDIA's CUBLAS



Highly optimized GEMV/SYMV kernels

NVIDIA has adopted for its CUBLAS 6.0 library



### "KAUST inside"



#### DEVELOPER ZONE

#### CUDA TOOLKIT DOCUMENTATION



#### CUDA Toolkit v6.0

#### **CUBLAS**

- 1. Introduction
- 1.1. Data layout
- 1.2. New and Legacy cuBLAS API
- 1.3. Example code
- 2. Using the cuBLAS API
- 2.1. General description
- 2.1.1. Error status
- 2.1.2. cuBLAS context
- 2.1.3. Thread Safety
- 2.1.4. Results reproducibility
- 2.1.5. Scalar Parameters
- 2.1.6. Parallelism with Streams
- 2.1.7. Batching Kernels
- 2.1.8. Cache configuration
- 2.1.9. Device API Library
- 2.2. cuBLAS Datatypes Reference
- 2.2.1. cublasHandle t
- 2.2.2. cublasStatus\_t
- 2.2.3. cublasOperation\_t
- 2.2.4. cublasFillMode t
- 2.2.5. cublasDiagType\_t
- 2.2.6. cublasSideMode\_t
- 2.2.7. cublasPointerMode\_t
- 2.2.8. cublasAtomicsMode t
- 2.3. cuBLAS Helper Function Reference
- 2.3.1. cublasCreate()
- 2.3.2. cublasDestroy()
- 2.3.3. cublasGetVersion()
- 2.3.4. cublasSetStream()

#### C. Acknowledgements

NVIDIA would like to thank the following individuals and institutions for their contributions:

- · Portions of the SGEMM, DGEMM, CGEMM and ZGEMM library routines were written by Vasily Volkov of the University of California.
- · Portions of the SGEMM, DGEMM and ZGEMM library routines were written by Davide Barbieri of the University of Rome Tor Vergata.
- Portions of the DGEMM and SGEMM library routines optimized for Fermi architecture were developed by the University of Tennessee.
   Subsequently, several other routines that are optimized for the Fermi architecture have been derived from these initial DGEMM and SGEMM implementations.
- The substantial optimizations of the STRSV, DTRSV, CTRSV and JTRSV library routines were developed by Jonathan Hogg of The Science
  and Technology, Lacricies Council (STFC). Subsequently, some optimizations of the STRSM, DTRSM, CTRSM and Excell have been
  served from these TRSV implementations.
- Substantial optimizations of the SYMV and HEMV library routines were developed by Ahmad Abdelfattah, David Keyes and Hatem Ltaief
  of King Abdullah University of Science and Technology (KAUST).

#### **Notices**

#### Notice

ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, "MATERIALS") ARE BEING PROVIDED "AS IS." NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE.

Information furnished is believed to be accurate and reliable. However, NVIDIA Corporation assumes no responsibility for the consequences of use of such information or for any infringement of patents or other rights of third parties that may result from its use. No license is granted by implication of otherwise under any patent rights of NVIDIA Corporation. Specifications mentioned in this publication are subject to change without notice. This publication supersedes and replaces all other information previously supplied. NVIDIA Corporation products are not authorized as critical components in life support devices or systems without express written approval of NVIDIA Corporation.

#### **Trademarks**

NVIDIA and the NVIDIA logo are trademarks or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

#### Copyright

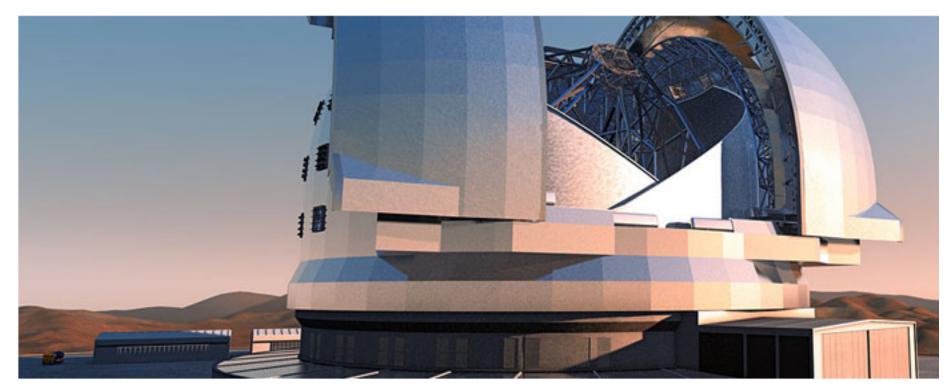
© 2007-2014 NVIDIA Corporation. All rights reserved.

This product includes software developed by the Syncro Soft SRL (http://www.sync.ro/).

## **KAUST** in European telescope (ELT)

### The European Extremely Large Telescope

The world's biggest eye on the sky



© ESO, https://www.eso.org/sci/facilities/develop/ao/ao\_modes.html

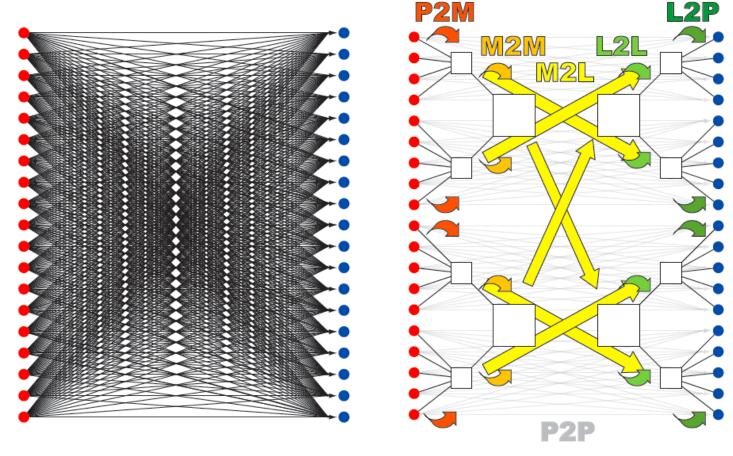


c/o Ahmad Abdelfattah & Ali Charara (KAUST), SC'14

# "Fast Multipole inside"

- Fast Multipole is an optimal algorithm for a very frequently used specialized kernel in computational science (the "potential solve")
- Fast Multipole, which has led to Gordon Bell prizes in astrophysics, molecular dynamics, materials science, arterial dynamics, etc. has been highly optimized by KAUST Research Scientist Rio Yokota in "ExaFMM"
  - Can we leverage its excellent architectural properties for more general operators in applied science?

## Fast Multipole communication schema



(a) Direct method

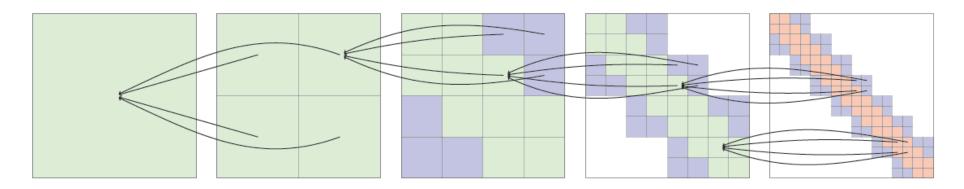
(b) Fast Multipole Method



c/o R. Yokota, KAUST, et al.

# "Algebraic Fast Multipole" (AFM)

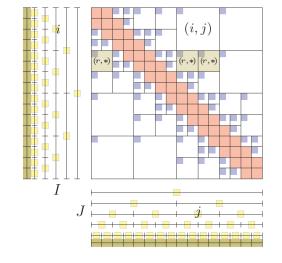
• We have cast general matrix operations (add, multiply, invert, etc.) in terms of the fast multipole recursive "tree-based" data structure



• We first significantly compress the matrix, saving storage and operations

May offer breakthrough in application performance

c/o Wajih Boukaram (KAUST)







### **Model for ECRC enablements**

Mission: enable scientists and engineers to take advantage of petascale hardware by overcoming the scalability bottlenecks of traditional solvers, and assist users to move beyond one-off "forward" simulations to inversion, optimization, sensitivity, UQ, validation (\$32M/10 years, 2001-11, PI: D. Keyes)













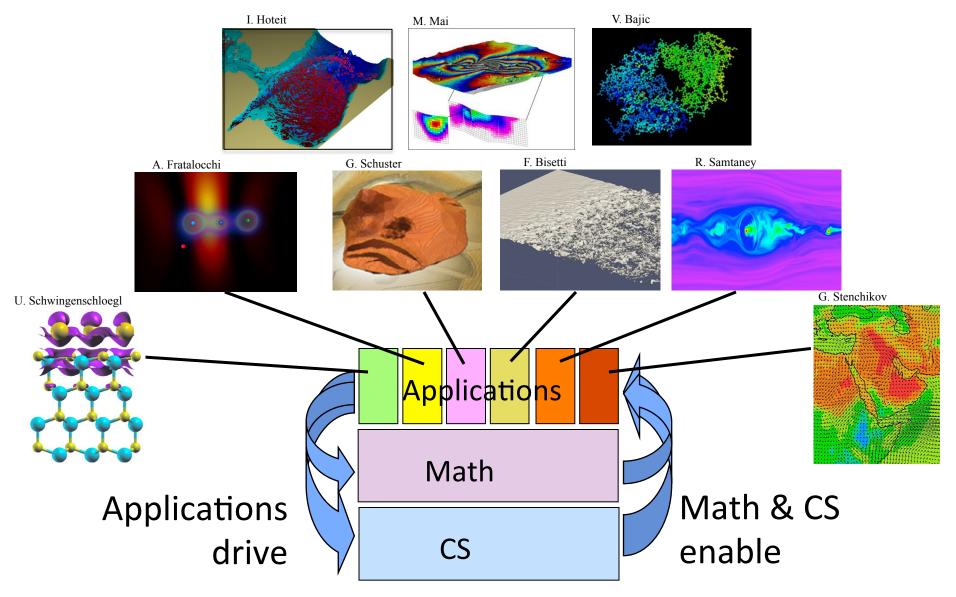




University of California at

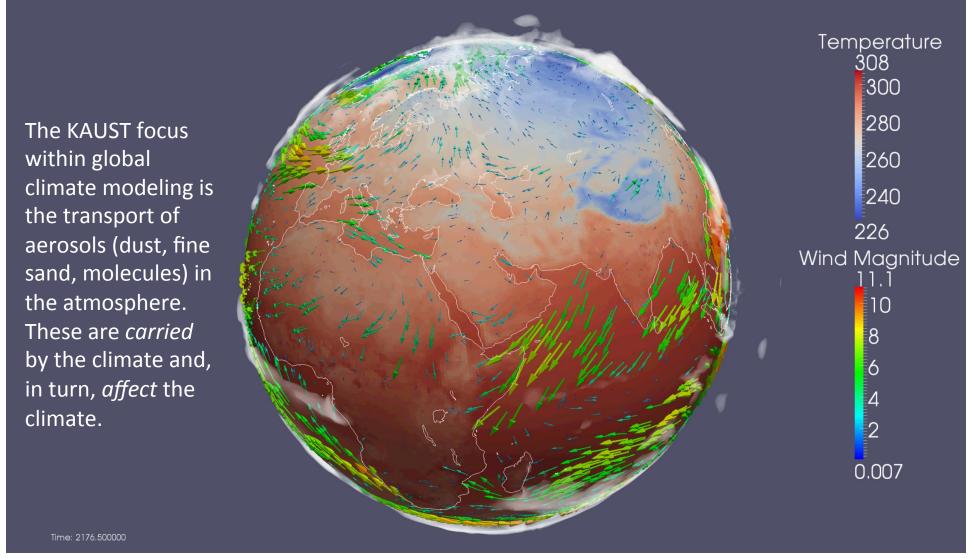
San Diego

### SciDAC-like model for software investments



# **Professor Georgiy Stenchikov Global climate simulation**

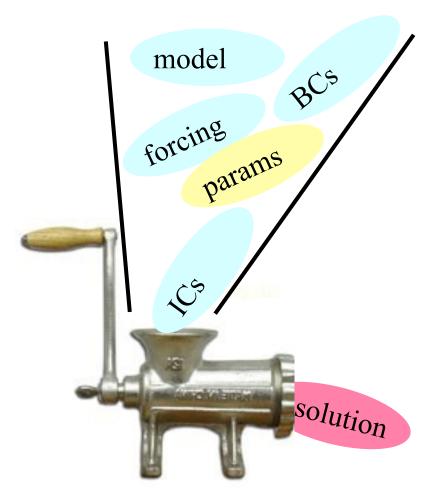


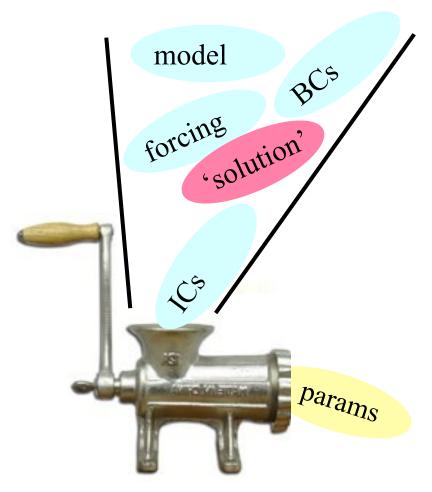


# Combining the 3<sup>rd</sup> and 4<sup>th</sup> paradigms: forward vs. inverse problems

forward problem

inverse problem

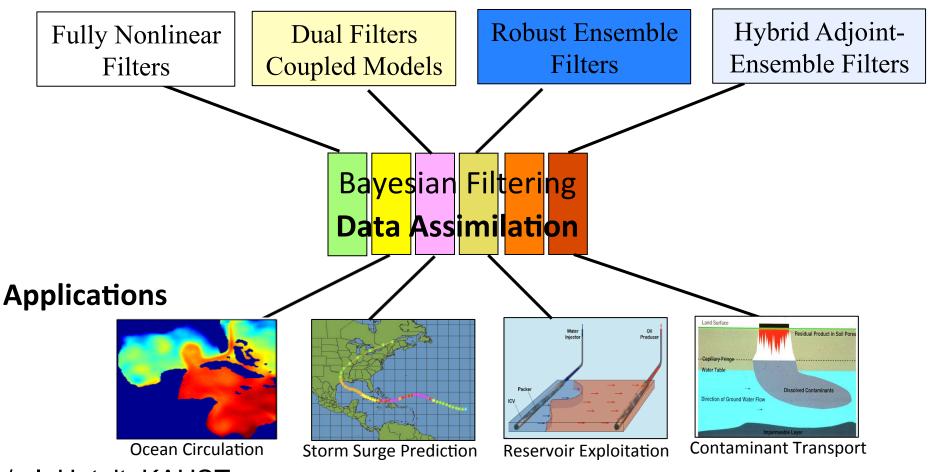




+ regularization

# Combining the 3<sup>rd</sup>and 4<sup>th</sup> paradigms: data assimilation

### **Theory**



c/o I. Hoteit, KAUST



- Wayne Gretzsky

Thanks to IESP and BDEC for attempting to ferret out and forge a consensus for productive research directions, and predicting where the puck is going to be.

