Computational Modeling and Analysis For Complex Systems
NSF Expedition in Computing

CMACS: An Overview

Edmund M. Clarke, Lead PI

Carnegie Mellon University

http://cmacs.cs.cmu.edu/

2nd Year Review Meeting, Carnegie Mellon University
November 3, 2011

Talk Outline

- Vision and Goals
- Challenges
- Research Team
- Accomplishments to date
  - Fundamental cross-cutting themes
  - Results on the Challenge Problems
- Education and Outreach
- Collaboration
- Further Work

Our Vision

To gain fundamental new insights into the **emergent behaviors** of complex biological and embedded systems through the use of **revolutionary**, highly **scalable**, and fully **automated** modeling and analysis techniques.

Our Goals

- **Scientific:** Develop Next-Generation Model Checking and Abstract Interpretation – MCAI 2.0
- **Societal:** Apply MCAI 2.0 to Challenge Problems in complex biological and embedded systems
- **Education & Outreach**: Build a program that
  - supports CMACS' vision of research and knowledge transfer
  - serves as a primary recruitment mechanism for students, especially those from under-represented groups

Model Checking

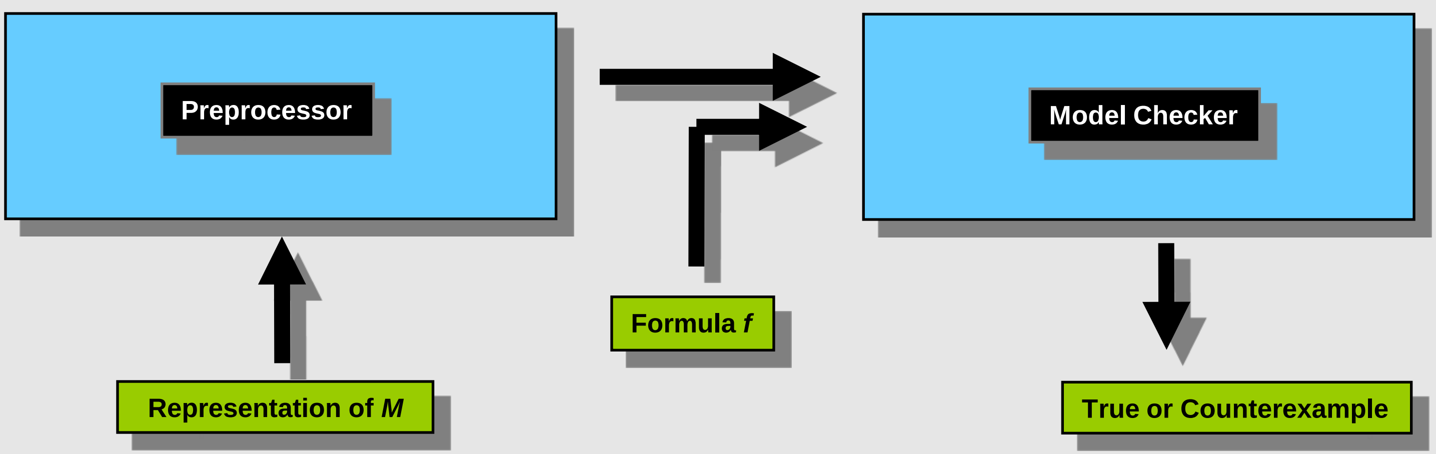
The **Model Checking Problem** (**Clarke**, Emerson, Sifakis ‘81):

Let ***M*** be a **state-transition graph**

Let ***f*** be a **formula of temporal logic**

e.g., a U b means “**a** holds true **U**ntil **b** becomes true”

Does ***f*** hold along all paths that start at initial state of ***M*** ?



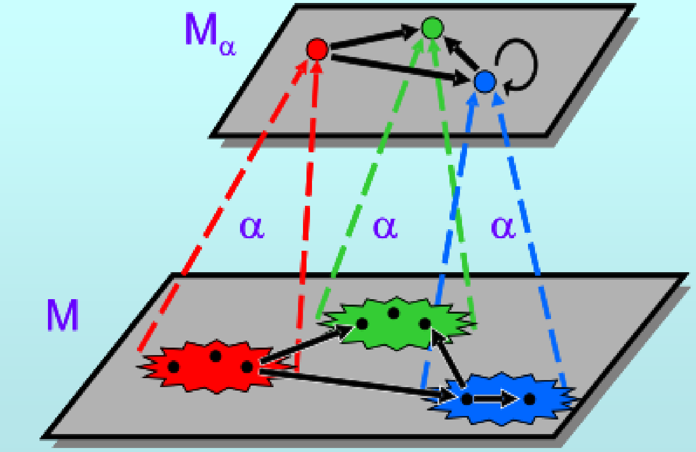
Many Industrial Successes

- Try 4195835 – 4195835 / 3145727 * 3145727.
  - In 94’ Pentium, it doesn’t return 0, but 256.
- Intel uses the SRT algorithm for floating-point division. Five entries in the lookup table are missing.
- Cost: $500 million
- Xudong Zhao’s Thesis on Word-Level Model Checking

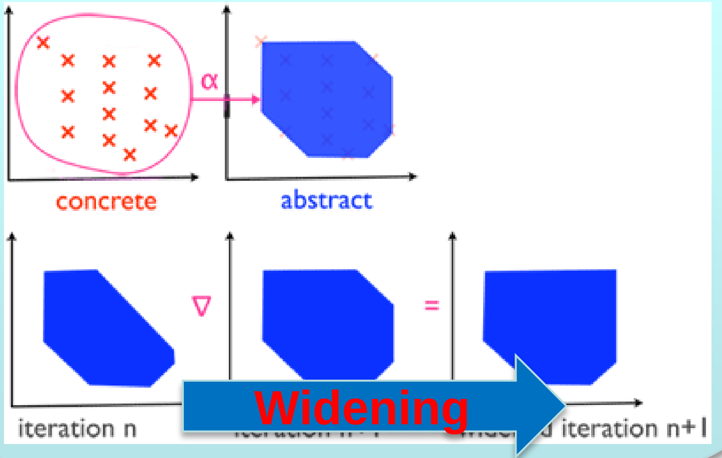
Abstract Interpretation

- Abstracts the **concrete semantics** of a system into a simpler **abstract semantics**
- Developed by **Cousot** & Cousot in 1977

Control Abstraction



Data Abstraction



Features of Abstract Interpretation

- **Automatic extraction of correct information** about the possible executions of complex systems
- Can be used to reason about **infinite state systems**
- **Scalability!** e.g., **A380** primary flight control system:
  - 1 million lines of C code
  - 34 hours to analyze
  - Numerous runtime errors were found statically and repaired
  - 0 false positives

CMACS …

- Is **rethinking** and developing an **integration** of Model Checking and Abstract Interpretation

- Is driven by the **centrality of computational modeling** in science & engineering

- Focuses on complex **biological** and **embedded** systems

- Is **cross-pollinating**: same techniques applicable in one domain **transfer to the other** (and beyond!)

Talk Outline

- Vision and Goals
- Challenges
- Research Team
- Accomplishments to date
  - Fundamental cross-cutting themes
  - Results on the Challenge Problems
- Education and Outreach
- Collaboration
- Further Work

Primary Challenge: Scalability

Key Scalability Issues:

Spatial Distribution

Stochastic Behavior

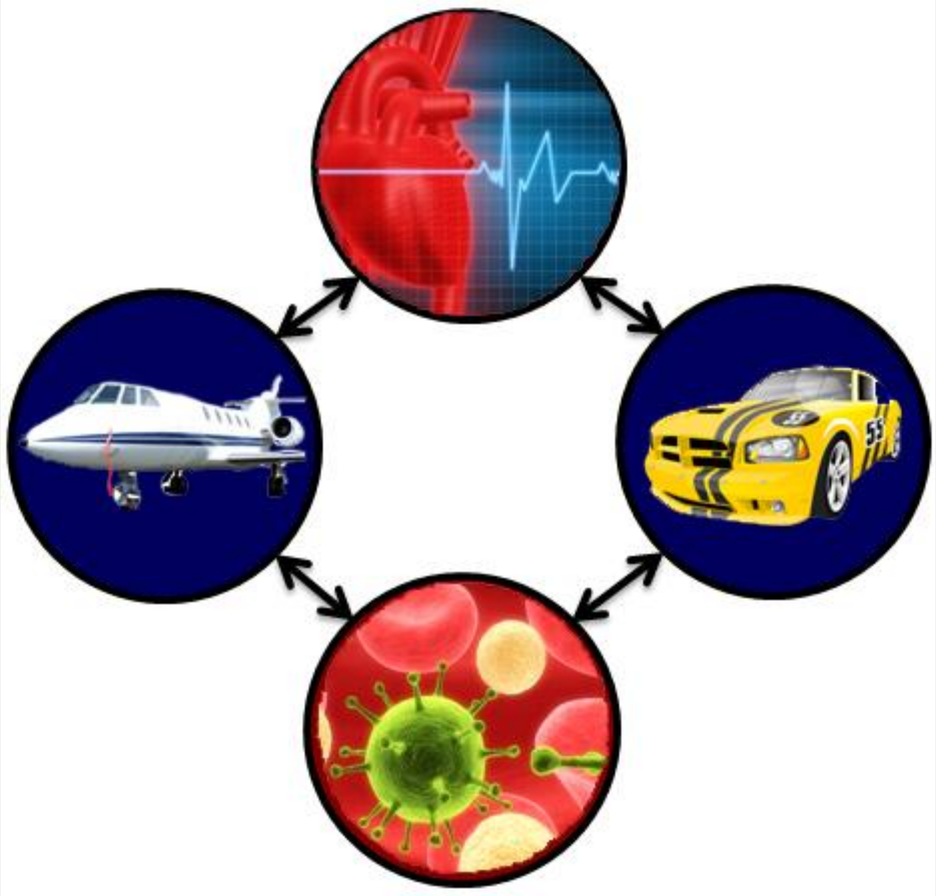
Highly Nonlinear Behavior

Mixed (Hybrid) Continuous-Discrete Behavior

Vast Numbers of System State Variables & Components

Complex Biological & Embedded Systems can exhibit any combination of these features

Challenge Problems



Systems Biology

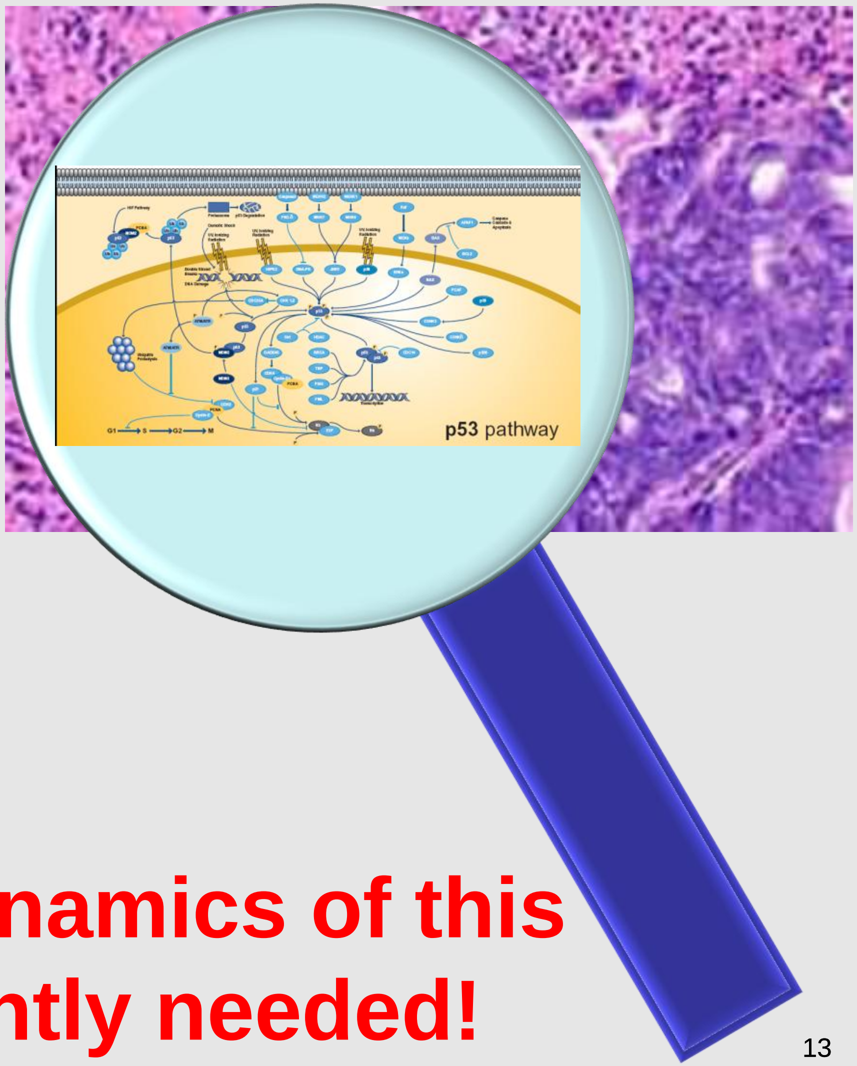
- Pancreatic Cancer
- Atrial Fibrillation

Embedded Systems

- Distributed Automotive Control
- Aerospace Flight Software

Pancreatic Cancer

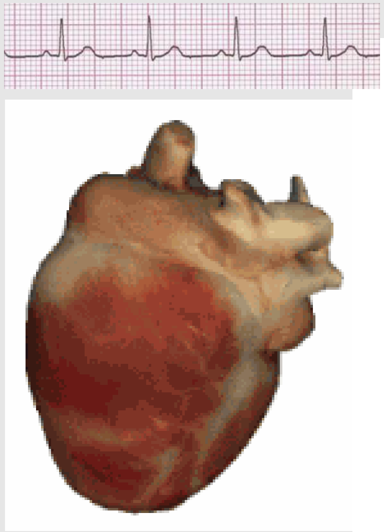
- 4th leading cause of cancer death in the US and Europe
- Five-year survival rate is only 4%
- Very few animal models
- Need computational models
- Building new analysis and verification tools



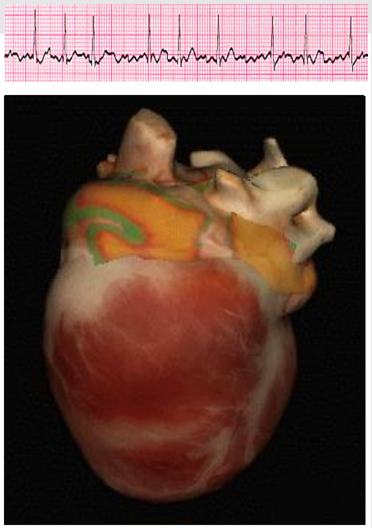
New insights into the dynamics of this deadly disease are urgently needed!

Atrial Fibrillation

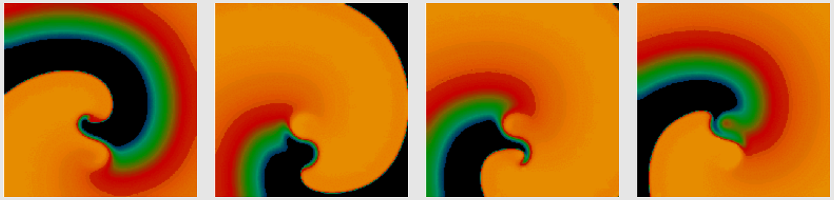
- **Most commonly diagnosed cardiac arrhythmia**
- **10 million Americans projected to have AF by 2050**
- **MCAI 2.0** can yield **reduced models** with virtually the **same dynamics**



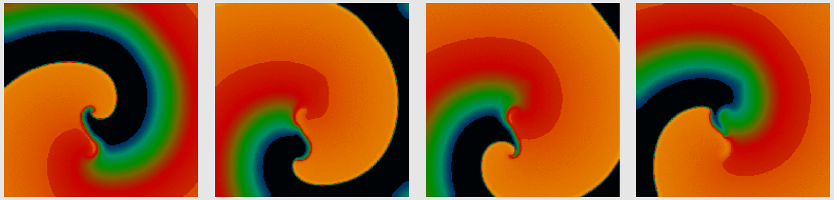
ECG SIMULATION



Normal Rhythm **Atrial Fibrillation**

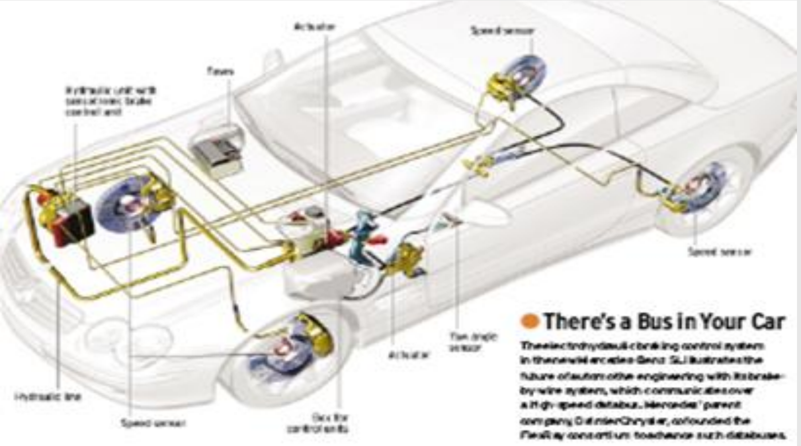


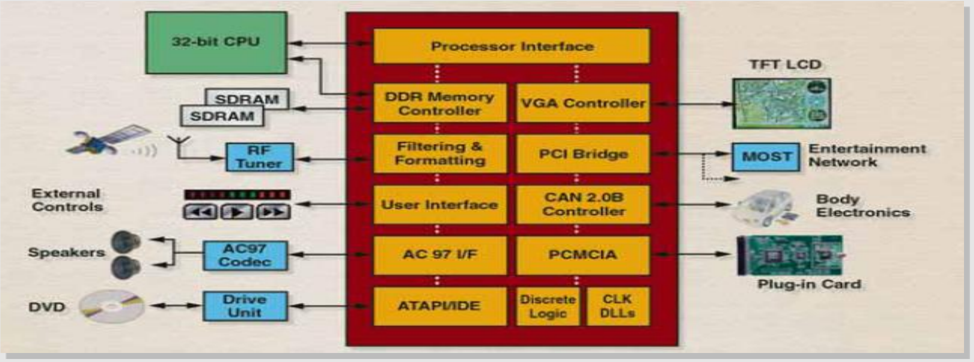
Full Model*: 4 snapshots during one period of the Ten Tusscher et al. model (17 variables)*

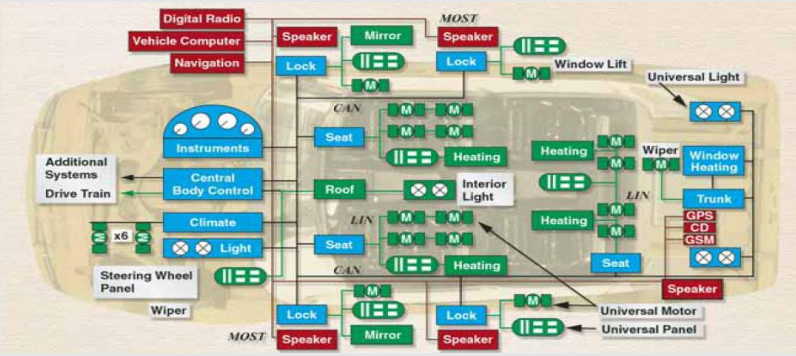


Reduced Model*: 4 snapshots during one period of the Bueno et al. model (4 variables)*

Automotive Embedded Systems







Do you trust your car?

Aerospace Systems: Software Driven!

Mars Polar Lander (1999)
landing-logic error

Mission Loss

Spirit Mars Rover (2004)
file-system error

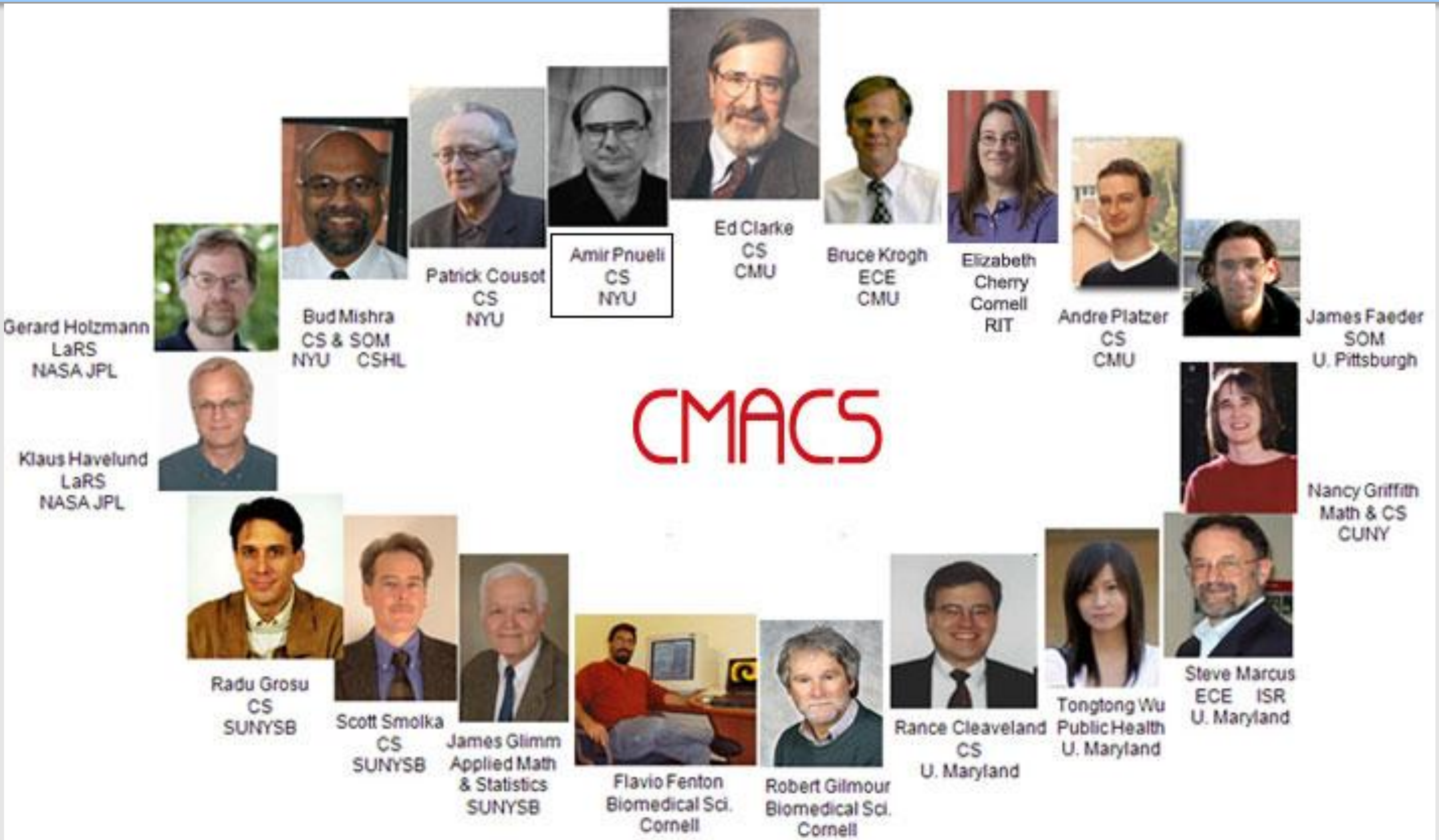
Airbus A380 Flight Deck

Do you trust flight software?

Talk Outline

- Vision and Goals
- Challenges
- Research Team
- Accomplishments to date
  - Fundamental cross-cutting themes
  - Results on the Challenge Problems
- Education and Outreach
- Collaboration
- Further Work

CMACS: Research Team



Team Member Highlights

- **Edmund Clarke:** co-inventor of Model Checking, co-recipient of 2007 ACM Turing Award and 1998 ACM Paris Kanellakis Theory and Practice Award, member of National Academy of Engineering
- **Amir Pnueli*:** recipient of the 1996 ACM Turing Award for introducing temporal logic into computer science, many honorary degrees
- **Patrick Cousot:** co-inventor of Abstract Interpretation, received 2008 Humboldt Research Award, 1999 Laureate of the CNRS silver medal, 2006 EADS Scientific Grand Prix
- **Gerard Holzmann:** recipient of the 2001 ACM Software System Award and 2006 ACM Paris Kanellakis Theory and Practice Award, member of National Academy of Engineering
- **Jim Glimm:** awarded 2002 National Medal of Science for his work in shock wave theory & other cross-disciplinary fields in mathematical physics, member of National Academy of Sciences

*The CMACS team sorely misses Amir Pnueli, who passed away on November 2, 2009.

Talk Outline

- Vision and Goals
- Challenges
- Research Team
- Accomplishments to date
  - Fundamental cross-cutting themes
  - Results on the Challenge Problems
- Education and Outreach
- Collaboration
- Further Work

Significant Achievements & Impacts

- New computational methods for **cancer**
- New computational methods for **cardiac dynamics**
- New automated modeling and verification techniques for **complex embedded systems**
- Highly successful 2010 and 2011 **Undergraduate Workshops** on Pancreatic Cancer and Atrial Fibrillation for students from urban, minority-serving institutions
- CMACS **Embedded Systems Industry** Workshop (20 Oct. 2011, CMU)

Cross-cutting Research Themes

- Statistical Model Checking
- Advances in Abstract Interpretation
- GPU-based real-time simulation of dynamical systems
- New breakthrough techniques for the analysis of hybrid systems

Statistical Model Checking

- **Systems Biology**
  - BioNetGen: verification of rule-based models of signaling pathways (BMC Bioinformatics ’10)
  - BooleanNet: verification of logical models of signaling pathways - probabilistic Boolean networks (in progress)

- **Embedded Systems**
  - Verification of time-bounded properties for stochastic Stateflow/Simulink models (HSCC ’10 + submitted)

- **PRISMATIC**
  - With Oxford and SIFT: added Statistical MC capabilities (and more!) to PRISM [funded by META II – DARPA]

Abstract Interpretation

Significant advances on **infinite-state** systems

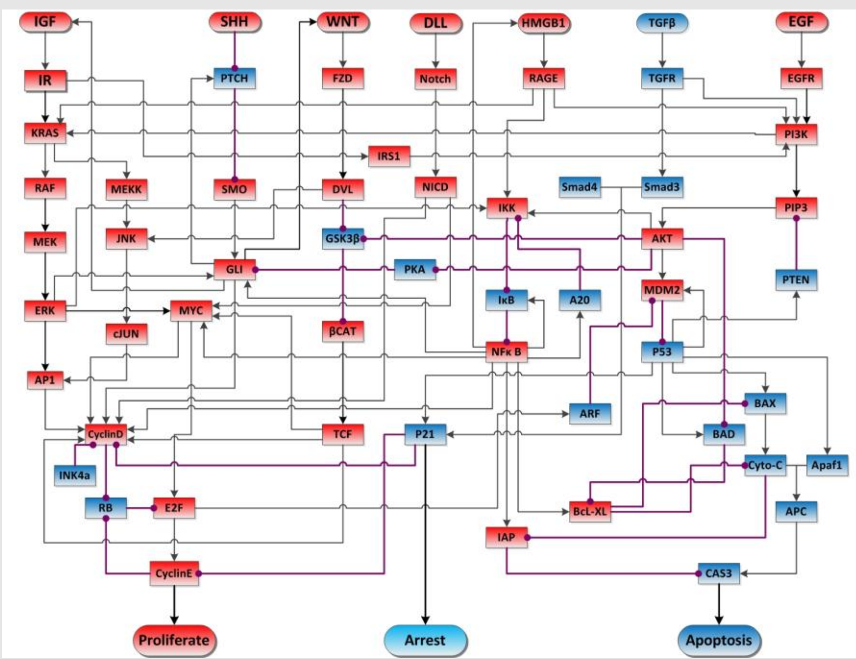
- Under-approximation — sufficient conditions for failure, i.e., necessary for success (VMCAI ’11)
- Analysis of array content (POPL ’11)
- Proof of termination/eventuality (POPL ’12)
- Probabilistic abstraction (submitted)

Integrating MC & AI

- A **major goal** – work very much in progress
- Combination of algebraic and logical abstractions
  - static analysis combined with SMT solvers or theorem provers (FoSSaCS ’11, The Future of Software Engineering ’10)
  - paves the way for a unification of two visions that have so far developed largely independently

Understanding Pancreatic Cancer through Computational Models

- CMACS researchers from CMU, Pitt & UPMC developed models & automated techniques for analysis of dynamic behavior of key biochemical processes in pancreatic cancer (*e.g.*, T cell differentiation, apoptosis, *etc*.)
- Potential applications in understanding the evolution of pancreatic cancer, and in drug design



Computational Model of PC Cell

Blue Nodes: tumor supressors
Red Nodes: oncoproteins/lipids
$\longrightarrow$: activation
$\multimap$: inhibition

Transcriptome Analysis for Pancreatic Cancer Survival

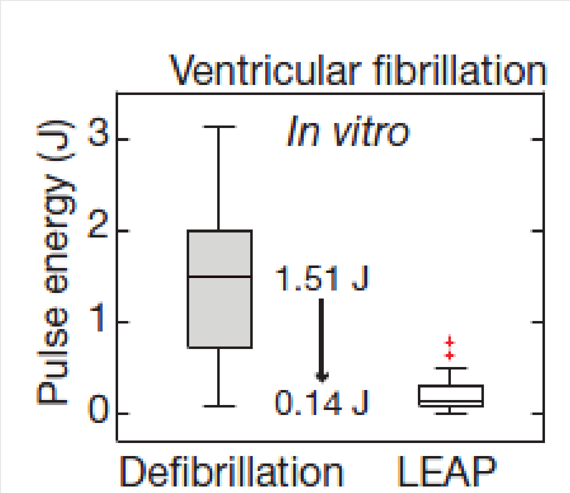
- Tongtong Wu, Haijun Gong, and Edmund Clarke have identified an **12-gene signature** for **pancreatic cancer survival** out of 43,376 candidate genes through Lasso-penalized Cox regression
- No previous studies on gene signatures that are directly related to pancreatic cancer survival
- **8 confirmed to be cancer-related** in the literature

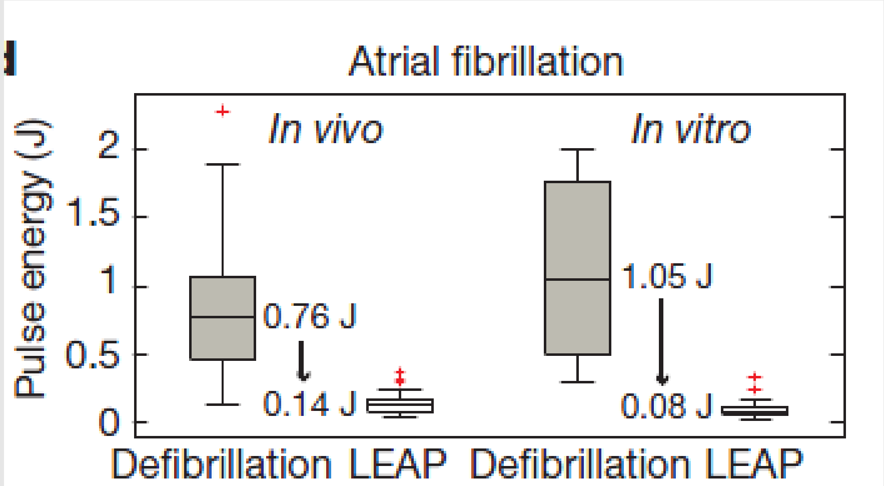
Gene	Function
RPS13	Promote cell cycle transition from G1 to S
PCYT1B	Regulates phosphatidylcholine biosynthesis
TREX2	Proapoptotic tumor suppressor, maintain the genomic integrity
ZNF233	Zinc nger protein, deregulated in kidney and pancreatic cancer
ATPAF1	Regulate oxidative phosphorylation pathway
RIMS1	Down-regulated in multidrug resistance gastric carcinoma
SLC43A2	Overexpressed in adenocarcinomas and squamous cell carcinoma
NRAP	Up-regulated in human pancreatic cancer

- 4 unknown: SLC22A8, C4orf35, C6orf81, and C6orf58

Control and Termination of Arrhythmias with Low-Energy Defibrillation

Low Energy Defibrillation (LEAP) tested for VF in vitro and for AF in vitro and in vivo (canine hearts).





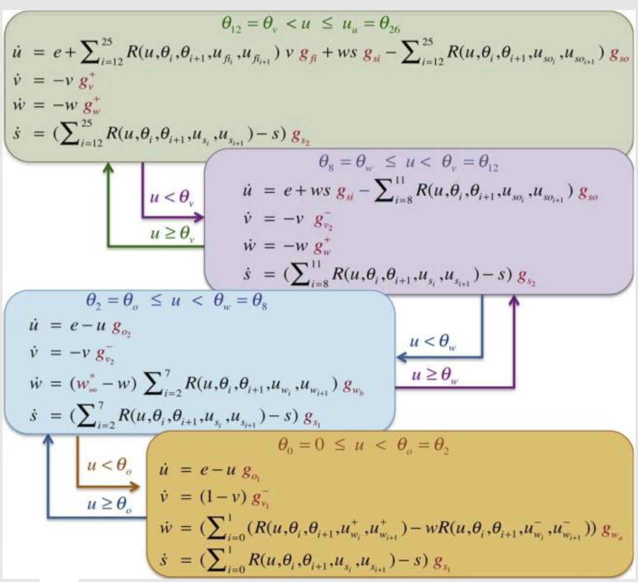
For Both AF and VF we have found successful defibrillation with LEAP using about 10% of the energy required by the standard 1 shock defibrillation protocol

Furthermore, using high resolution mCT We obtained detail vessel distribution of the heart and found a scaling law which was used to obtain a theory that explains the mechanism behind LEAP.

These results appeared this year in *Nature* 475: 235-239; 2011.

First Automated Formal Analysis of Realistic Cardiac Cell Model

- CMACS researchers from Stony Brook, Cornell & NYU succeeded in carrying out the first automated formal analysis of a realistic cardiac cell model
- Determined parameter ranges that lead to **loss of excitability**, a precursor to e.g. ventricular fibrillation

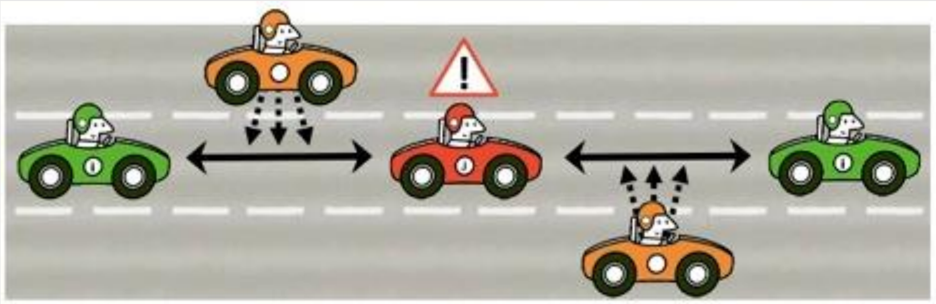


Multiaffine Hybrid Automaton model of Fenton et al.'s Minimal Cardiac Cell model

Such automata commonly used in the analysis of Genetic Regulatory Networks

How to Avoid Bugs while Driving on the Highway

- André Platzer, Sarah Loos, and Ligia Nistor have developed a protocol for **distributed adaptive cruise control** for highway traffic.
- Has further developed verification technology with which he can **prove** that protocol will successfully prevent collisions



Automated cars driving on the highway

Talk Outline

- Vision and Goals
- Challenges
- Research Team
- Accomplishments to date
  - Fundamental cross-cutting themes
  - Results on the Challenge Problems
- Education and Outreach
- Collaboration
- Further Work

NSF-CMACS Annual Workshop Series

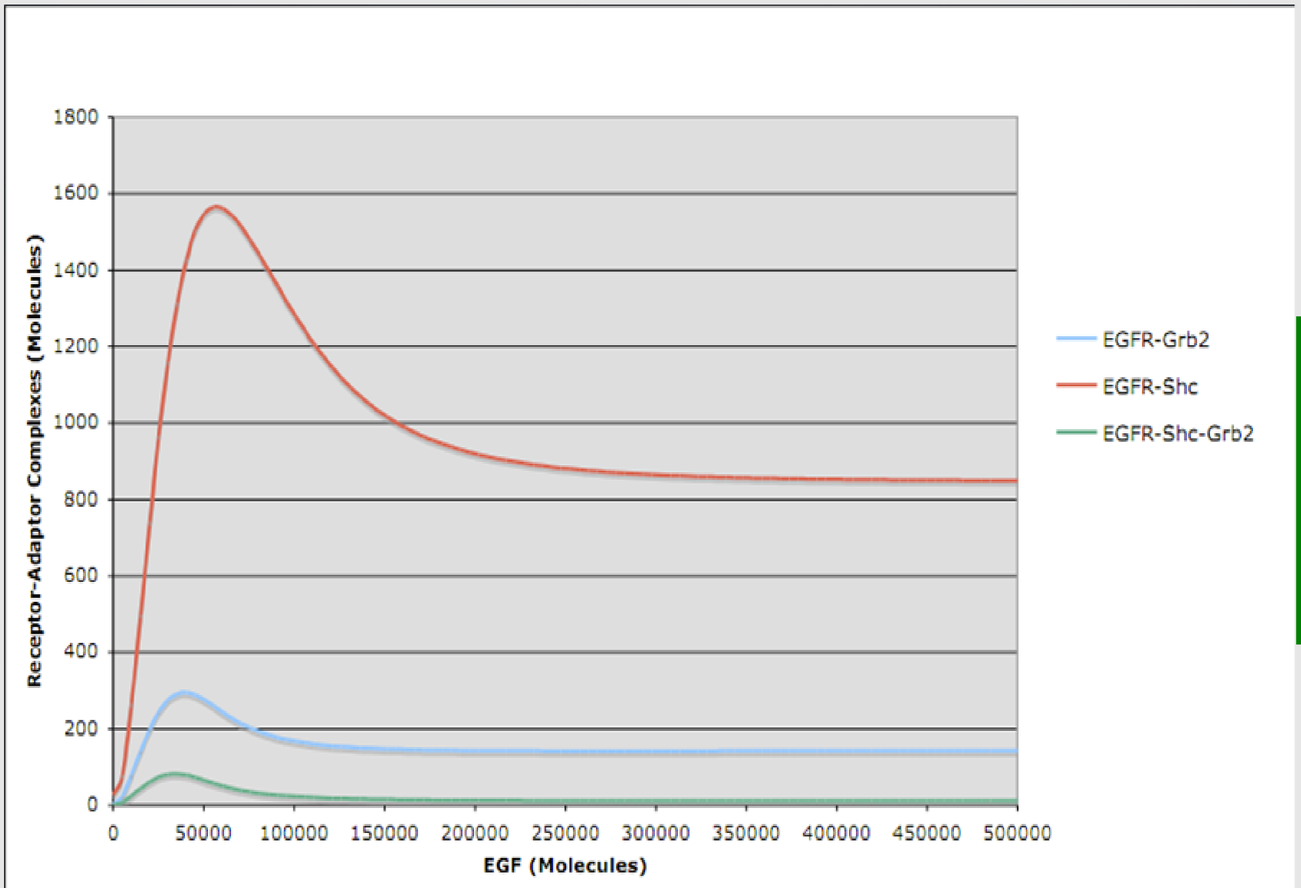
- Innovative educational program centered around annual workshops series which seeks to develop scientific interest & skills of students from urban, minority-serving institutions
  - **It would not have even been possible** without CMACS
- Each a **highly intensive 3-week workshop** held at Lehman College (part of CUNY) in the Bronx

Nancy Griffeth:
CMACS Educational Program Director

Flavio Fenton:
Program Co-Director

Jan 2010: Workshop on Pancreatic Cancer

- Focus on mathematical and computational tools for modeling biological systems, esp. **EGFR** receptor and its role in PC



By Ilya Korsunsky et al. Ilya now Junior Research Fellow in Bud Mishra's group

Jan 2011: Workshop on Atrial Fibrillation

- **Student co-authored** paper published in *Advances in Physiology Education*

Teaching In The Laboratory

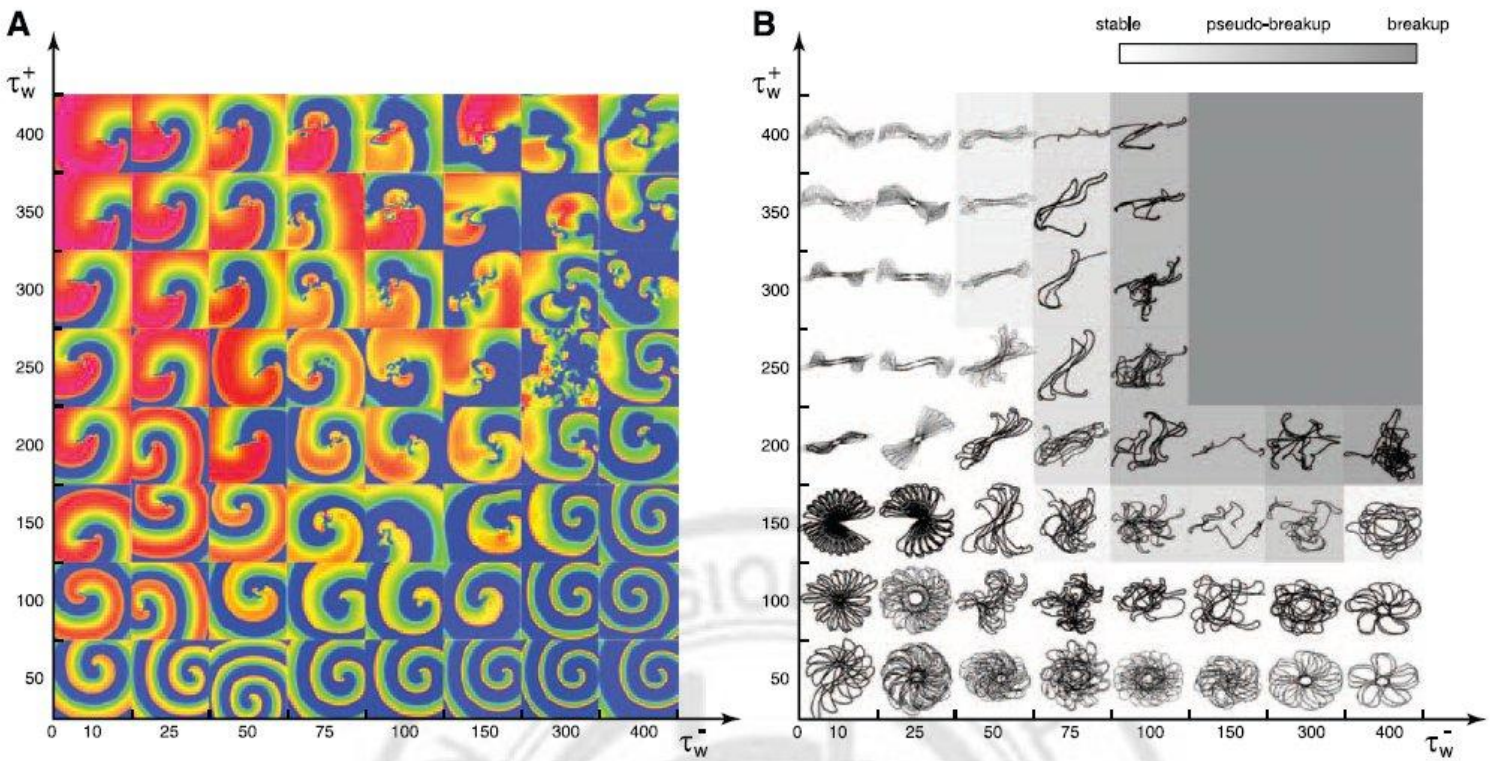


Fig. 5. Spiral wave snapshots (*A*) and tip trajectories (*B*) for combinations of the following parameters: the time constant to inactivate the slow inward Ca^{2+} current (τ_w^+) and the time constant to recover from the inactivation of the slow inward Ca^{2+} current (τ_w^-). Instability is promoted by increasing the values of both parameters, leading to quasi-breakup or full breakup (gray shaded regions in the tip trajectory plot). See text for details.

Talk Outline

- Vision and Goals
- Challenges
- Research Team
- Accomplishments to date
  - Fundamental cross-cutting themes
  - Results on the Challenge Problems
- Education and Outreach
- Collaboration
- Further Work

CMACS: Whole >> [Sum of Parts]

- Many breakthroughs due to **new**, **cross-institutional**, **cross-disciplinary** **collaborations**
- Typical example: Atrial Fibrillation Research

Stony Brook	Cornell/RIT	NYU
Bartocci (Computer Sci)	Cherry (Biomedical)	Le Guernic (Computer Sci)
Glimm (Applied Math)	Fenton (Physics)	
Grosu (Computer Sci)	Gilmour (Biomedical)	
Smolka (Computer Sci)		

CMACS: Whole >> [Sum of Parts]

- Another example: Pancreatic Cancer Research

CMU	Pitt	UMD	UPMC
Clarke (Computer Sci)	Faeder (Sys. Biol.)	Wu (Public Health)	Lotze (Cancer Inst.)
Gong (Computer Sci)	Miskov-Z. (Sys. Biol.)		
Wang (Computer Sci)			
Zuliani (Computer Sci)			

- June '11: **Translational Genomics Research Institute**
  - CMU group visited TGen (meeting Rich Posner and Daniel Von Hoff)

Talk Outline

- Vision and Goals
- Challenges
- Research Team
- Accomplishments to date
  - Fundamental cross-cutting themes
  - Results on the Challenge Problems
- Collaboration and Knowledge Transfer
- Education and Outreach
- Further Work

Future Work:
What Do the Next Three Years Hold?

- Discovery of **more detailed, realistic & probing computational models** of the biological & embedded systems we are so invested in studying

- Development of even **more efficient verification technology**, allowing us to tackle more expressive **properties** and more sophisticated **systems** (e.g. 2D & even 3D cell structures)

- More & wider cross-institutional & cross-disciplinary **collaborations**; e.g.
  - apply CMU statistical model checking to 2D & 3D cardiac models

- Education & Outreach: Winter Workshops at Lehman, plus Summer Workshops (at CMU?)