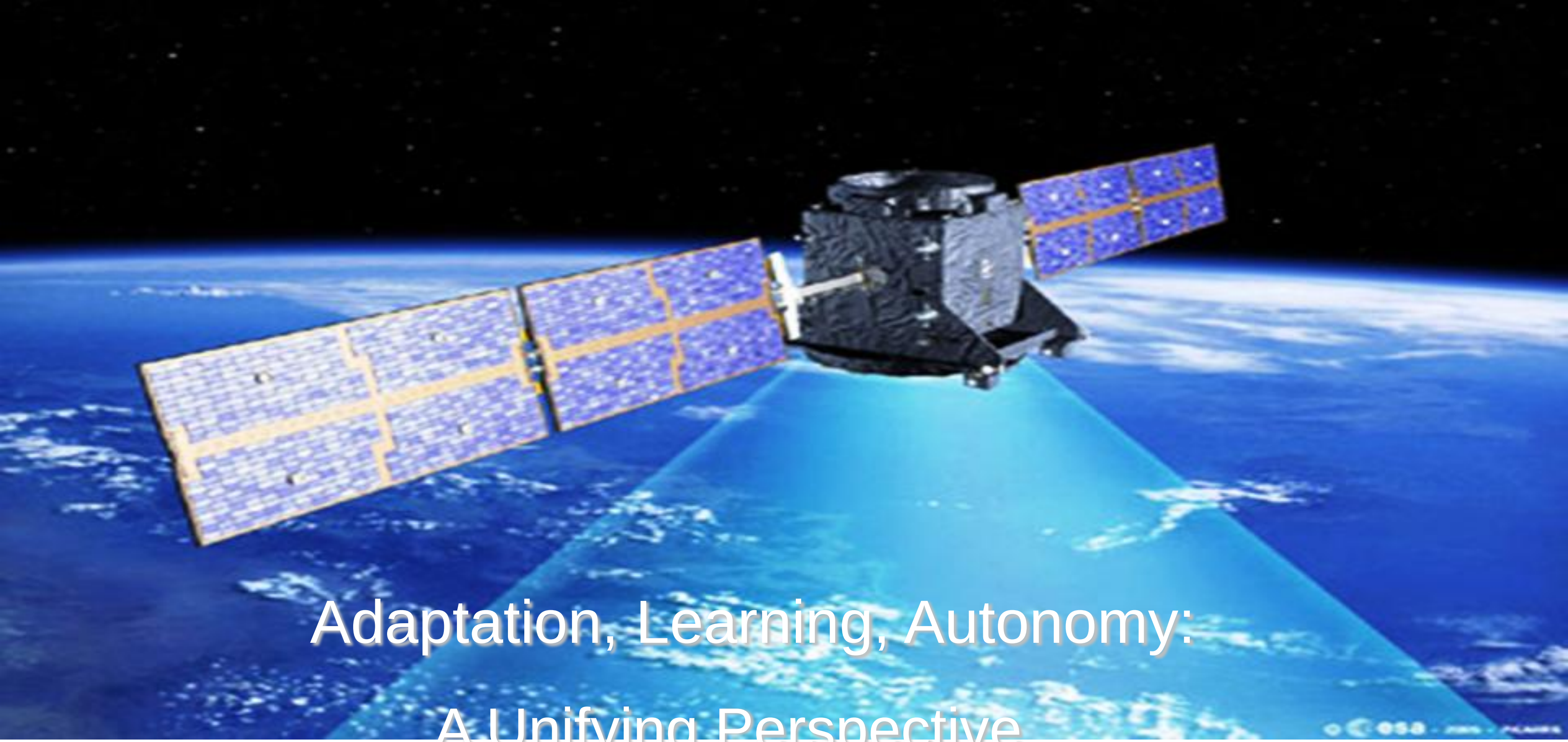


Adaptation, Learning, Autonomy:
A Unifying Perspective

Lena Valavani

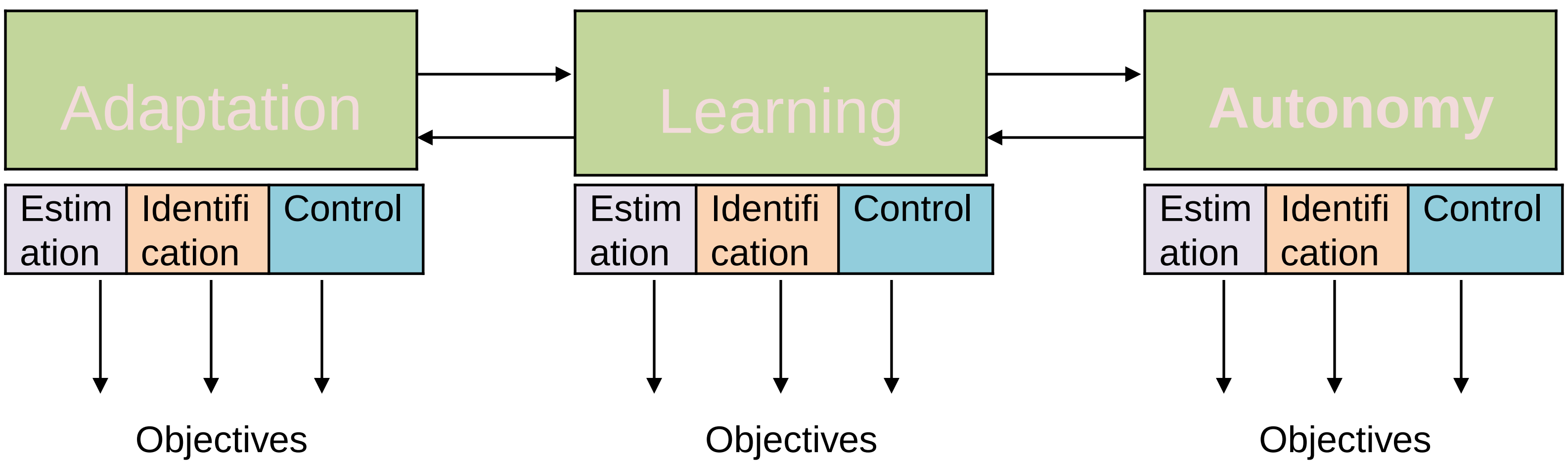
HSS

Hellenic Space Systems

(www.hellenicspacesystems.com)

LCCC, April 2010

Connections? Interdependencies?



Commonalities : Sharing of Basic Mathematical /Systems Concepts
Convergence to a ‘Set’

Differences : Objectives, Context

Adaptation, Learning, Autonomy

- Basic Synthesis Modules for Composite /Derivative Applications Areas (e. g. Cognition, etc)

Issues

- Performance vs Robustness vs Adaptivity
- Robustness vs Adaptivity

    - (Global) Robustness may require Adaptivity but

    - Too much Robustness also limits Adaptivity

Some Cognition Definitions

40 definitions of cognition & cognitive system from euCognition (http://www.eucognition.org/euCognition_2006-2008/definitions.htm)

- Cognition is the ability to relate perception and action in a meaningful way determined by experience,learning and memory. Mike Denham
- A cognitive system possesses the ability of self-reflection (or at least self-awareness). Horst Bischof
- Cognition is gaining knowledge through the senses. Majid Mermehdi
- Cognition is the ability to ground perceptions in concepts together with the ability to manipulate concepts in order to proceed toward goals. Christian Bauckhage
- An artificial cognitive system is a system that is able to perceive its surrounding environment with multiple sensors, merge this information, reason about it, learn from it and interact with the outside world. Barbara Caputo
- A machine is said to be a cognitive system if a human observer cannot detect a difference between the behavior of this machine and the behavior of an evolved animal (either human or animal relative). Michel Olivier
- Cognition is self-aware processing of information. Cecilio Angulo
- Cognitive systems are systems with knowledge-based behavior. Knowledge is agent-owned architecture-specific models of realities to interact with. Ricardo Sanz
- Cognitive Systems are ones that are able to extract and (most importantly) represent useful aspects of largely redundant, possibly irrelevant sensory information in a form that is most conducive to achieving a particular high level goal. Sethu Vijayakumar
- A cognitive system should be defined by its capability of finding the relationship of some or all the information coming from its inputs (sensors) in order to generate the best (internal or external) outputs after comparing this information with previously "memorized" input-output patterns. Boris Duran

- Cognition is the act of segmenting and recognizing a perceptual event and grounding (binding) it to a symbol (meaning). Axel Pinz

Adaptation, Learning, Autonomy

- Bottom Line : Stability is a basic requirement for information/process 'convergence'

Estimation	Identification	Control
-Observers	-ARMA	-STR
-Kalman Filters	-ARMAX	-MRAC
-Detection Filters	-RLS	-Gain Scheduling

etc.

In fact, an SPR or Passivity condition is ubiquitous (whether directly or indirectly realized)

Contraction Maps

- Stability and SPR/Passivity conditions are CM.
- Generalized concept, applicable to both Linear and Nonlinear systems (Lumped, Distributed)
- Under certain conditions, modular stability is also preserved in various combinations of contraction maps

Contraction Maps (cont'd)

- So, for the system

$$\dot{x} = f(x,t)$$

If f (x,t) defines contracting dynamics, with an appropriate *LTI* metric, then so does any *translated* and *scaled* version of it

$$f(a(t)x - b(t), t)$$

α(t) ,b(t) smooth, α(t) >0 uniformly so

Contraction Maps (cont'd)

Also,

- Parallel combinations of contractions, under the same (LTI) metric, also preserve the property

- Hierarchical combinations of contractions, with possibly different metrics, are also contractions

- Similarly for feedback combinations

Contraction Maps (cont'd)

- Contractions very useful for estimation, identification, control for linear and nonlinear systems, for distributed systems , etc.

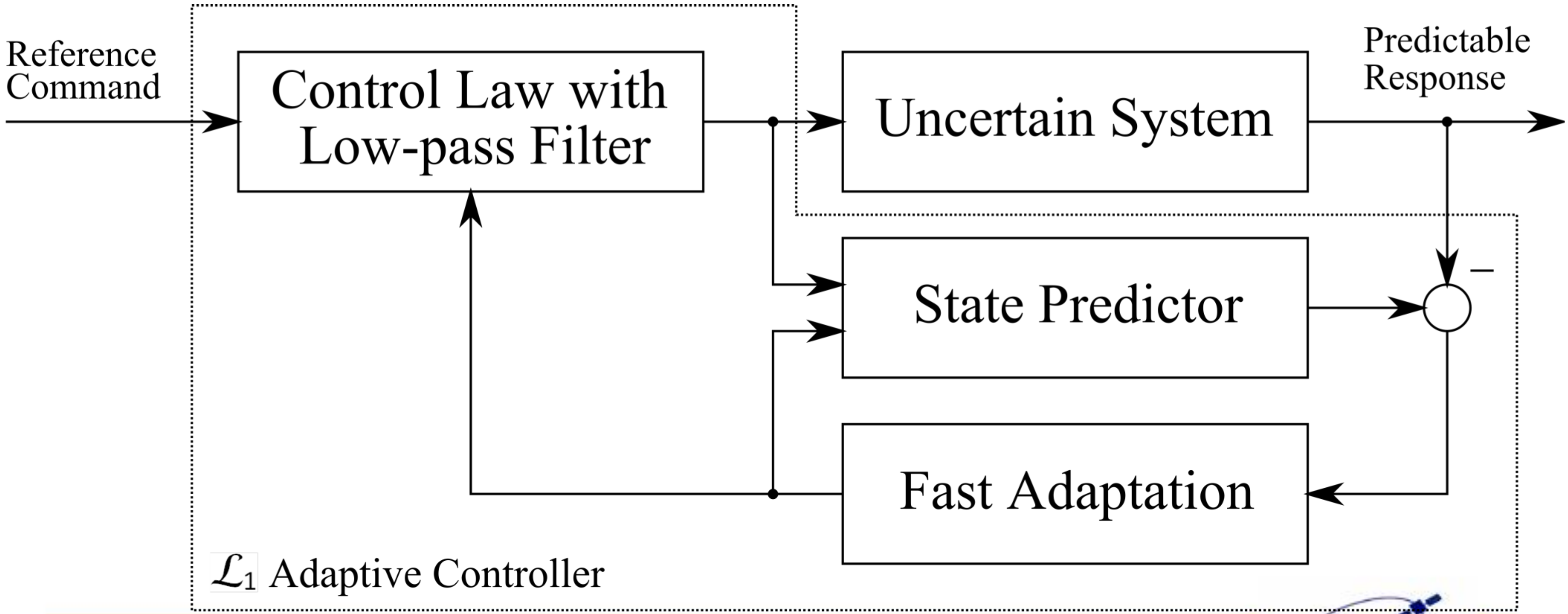
- Are they the answer ?
  - -mostly as analysis tool
  - -not a prescriptive synthesis tool; only insights
  - -restrictions apply

The Challenge

- Under the ‘same’ information , the key to success lies in picking the right *architecture* of system modules and their interconnections, with appropriate metrics
- No prescriptive ways of what specific architecture will work; importance of good understanding of the under-lying process and the design objectives
- Synthesis then is a challenge

Case in Point : L1 Adaptive Control

Architecture



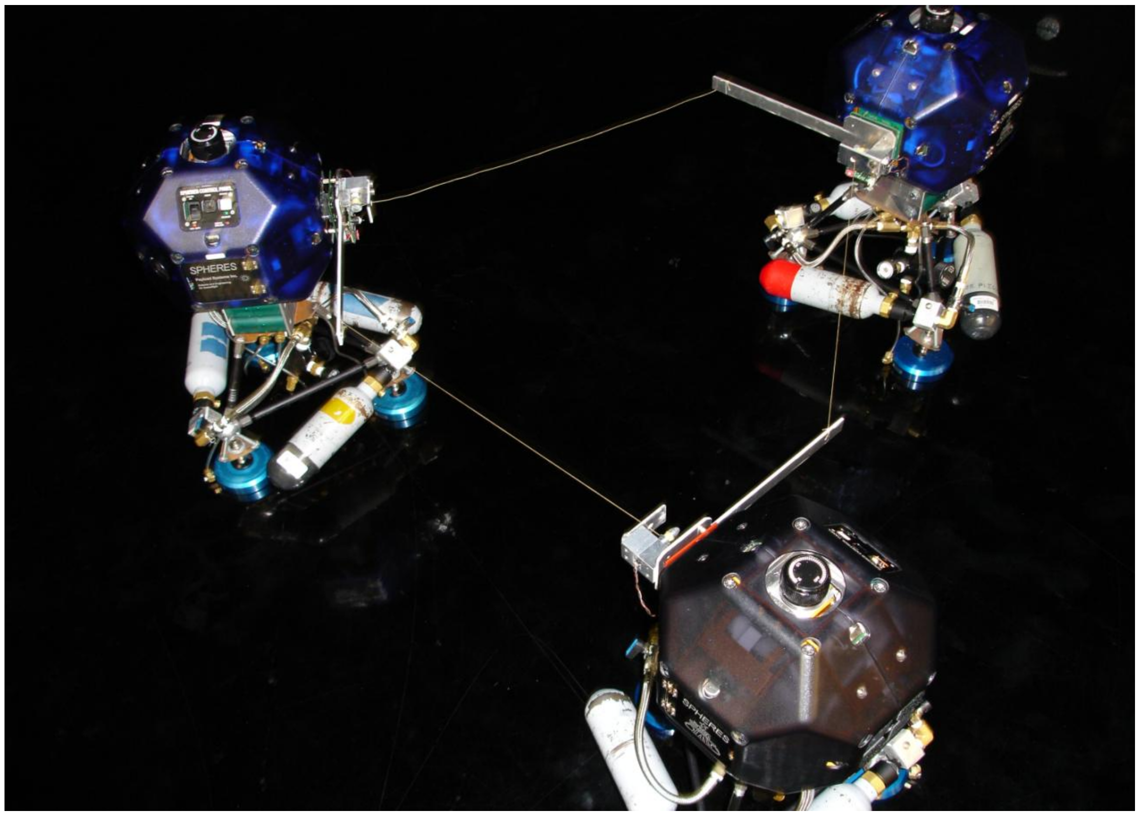
Courtesy of Prof. N. Hovakimyan

Some Interesting Distributed Systems

- Spacecraft Using Both GPS and Inter-Vehicle Ranging
- Formation Flying (Satellites, UAVs, etc.)
- Brain Research(Spatial Orientation, Vestibular response)
- SESAR/Next Gen Networks
- Earth Observation and a Multitude of Other Distributed Scenaria

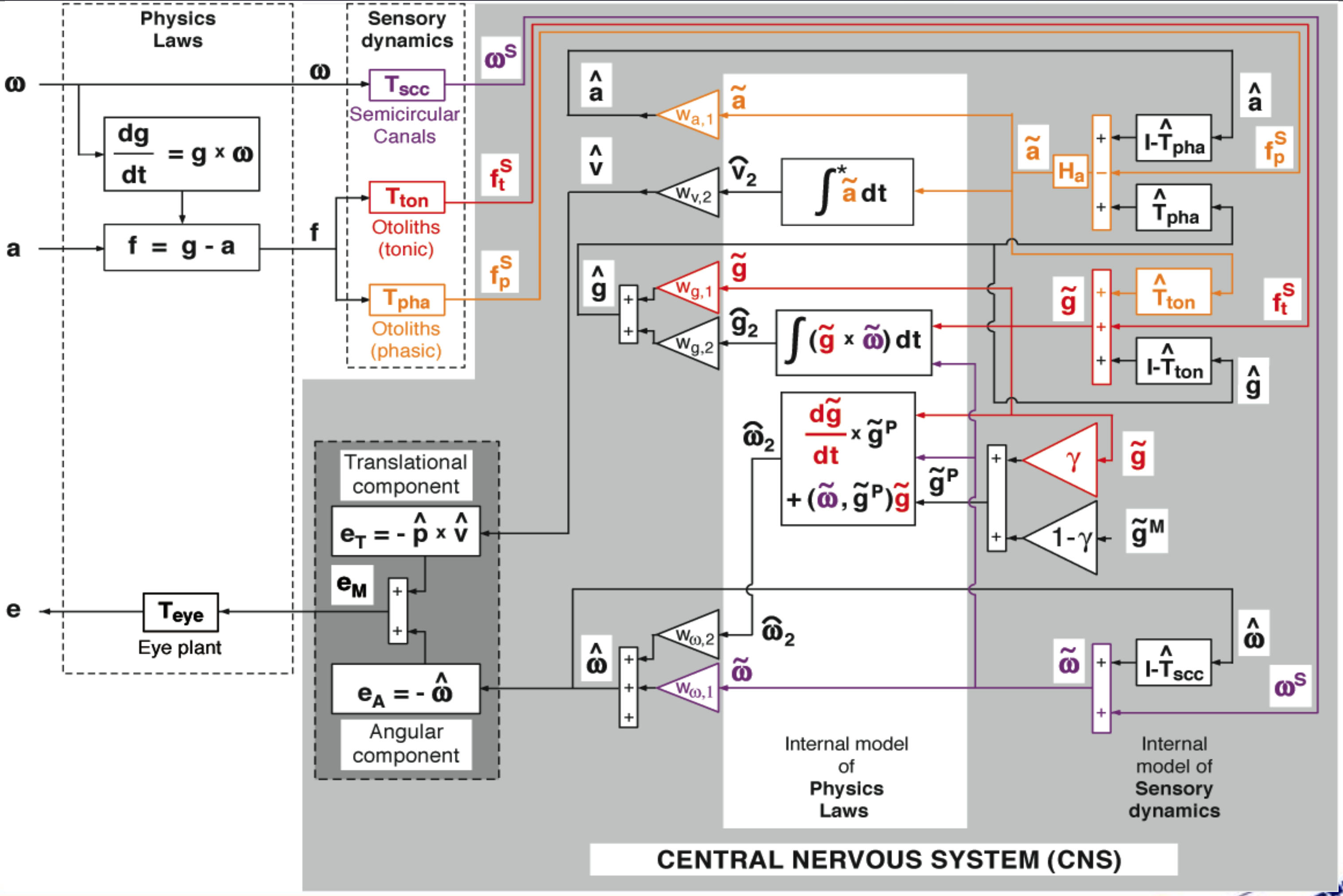
Examples of Distributed Systems / Networks

Formation Flying Satellites (SPHERES)



Courtesy of MIT's SSL (Astronaut G. Chamitoff)

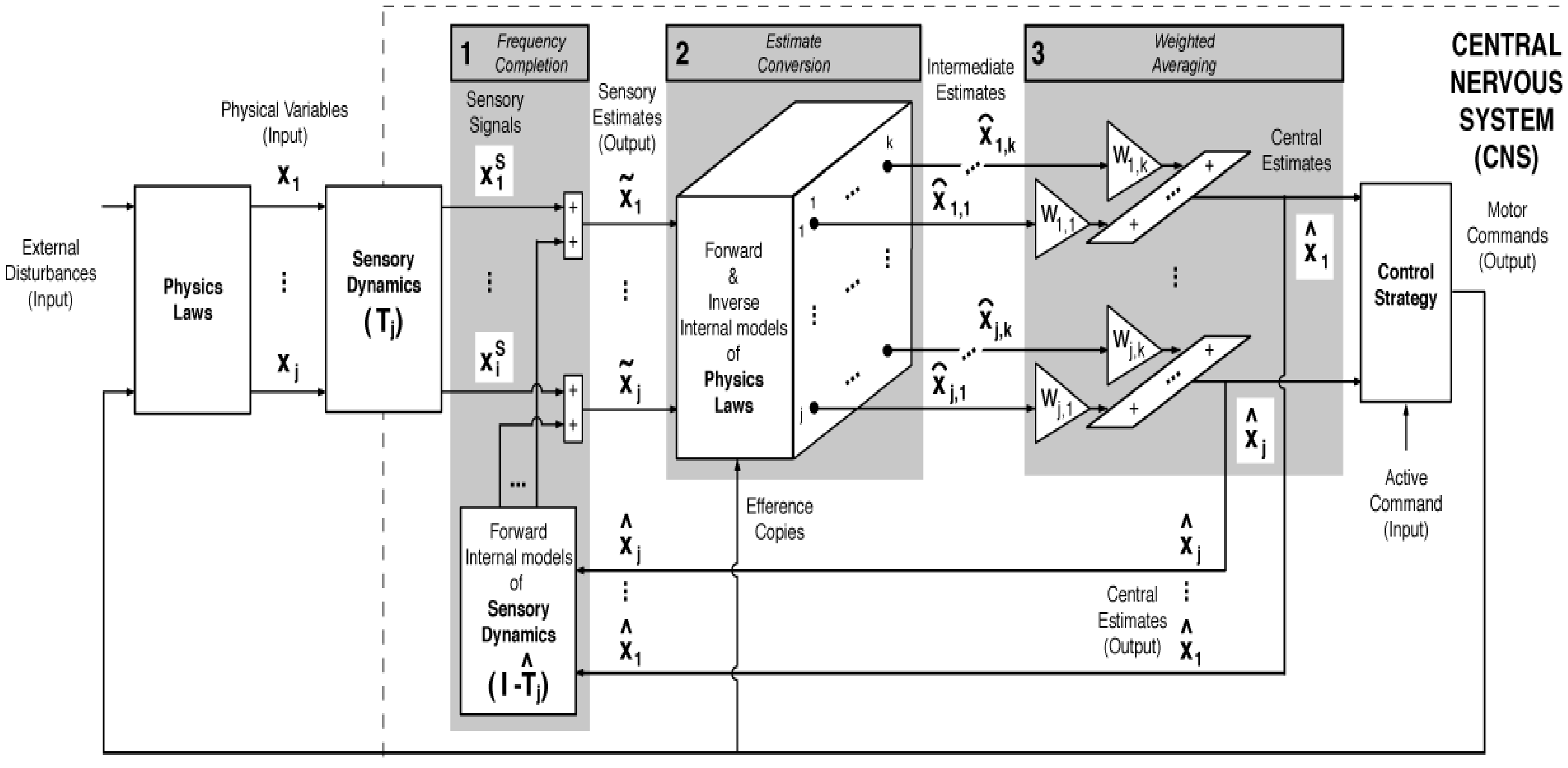
Distributed Network : The CNS



Jenks Vestibular Lab, MEEI

HSS

The CNS



Jenks Vestibular Lab, MEEI

The CNS

- Performs extensive parallel processing; has ‘collocated sensors and actuators (inherently suitable architecture for contraction type analyses)
- Carries out optimal information fusion, in a distributed nonlinear setting (without solving Riccati equations!)
- Neurons can be considered as ‘particles’: hence use particle filter formulations

Remark: As in all biological systems , in CNS also, inter-actions take place across (sub)systems and hierarchies, yet they remain modular

HSS

An Interesting Issue in Nonlinear Estimation

- All realistic scenaria admit to nonlinear propagation dynamics and measurements (EKF most popular)

- But when the measurements are (highly) nonlinear and wide ranging in accuracy(precise vs coarse measure-ments), filter divergence is certain to arise

- Almost all “Nonlinear” filters diverge:
    - The EKF and versions thereof
    - The GSOF,etc.
    - Various ‘Bump-Up’ Versions of Above

The Underlying Problem (cont'd)

- Even the UKF diverges, though specifically designed to mitigate 'linearization inaccuracies'!

Remark: Full fledged Particle Filters prohibitively computationally intensive for most (real-time) applications

- Hence , use simpler versions that can function in real time (formation flying requirements, 'correcting device' for neural disorders in spatial orientation)

The Crux of the Matter

- The propagation equations for all filters are accurate enough, despite individual differences
- The crux of the problem lies in the assumption that the state update is a linear function of the residuals (innovations).
- All filters, including the UKF utilize this assumption in their update mechanisms
- Key hypothesis: assumption may not hold for problems with high measurement nonlinearities

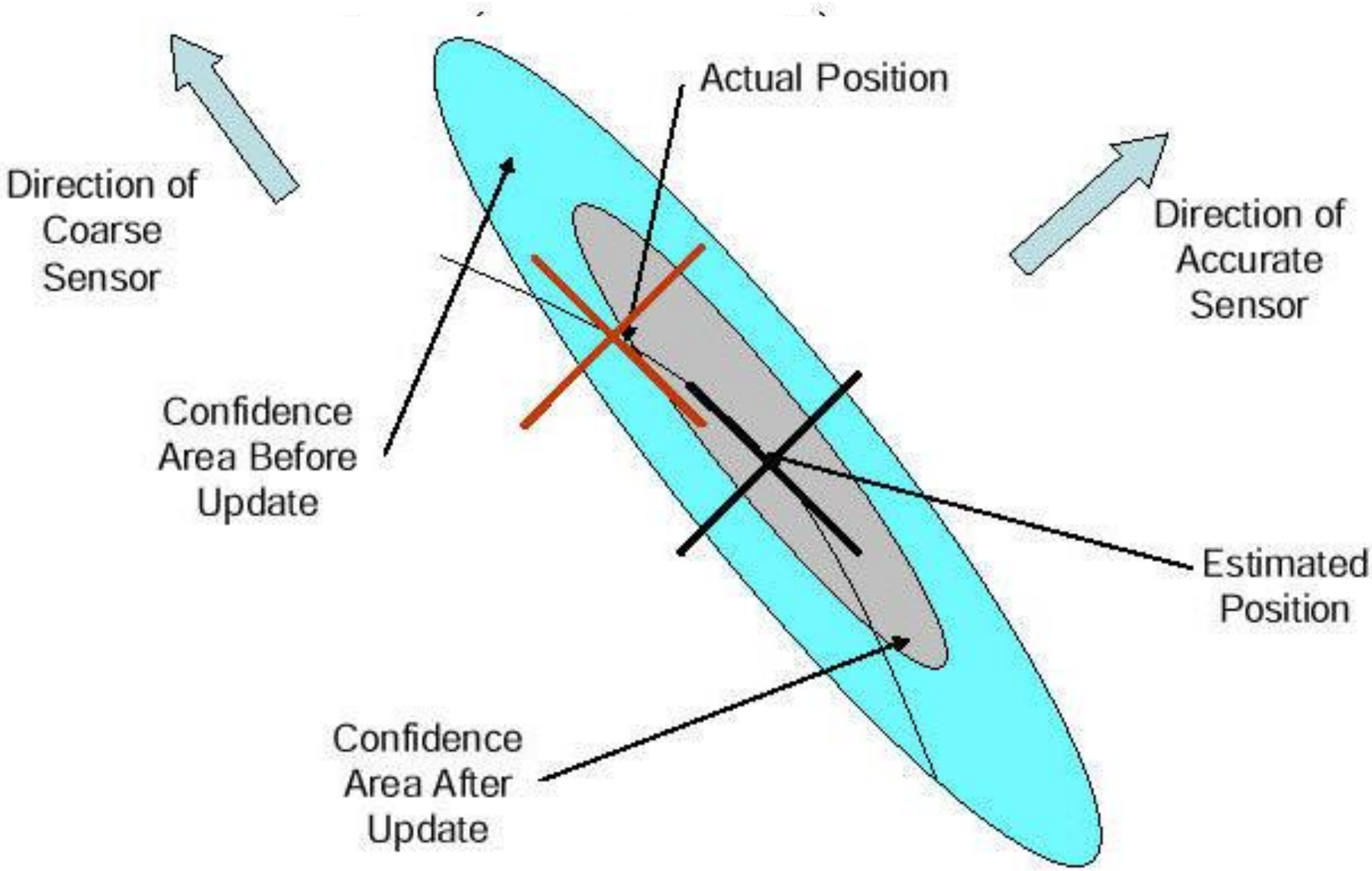
  - starting point that motivates our solution

The Divergence Mechanism

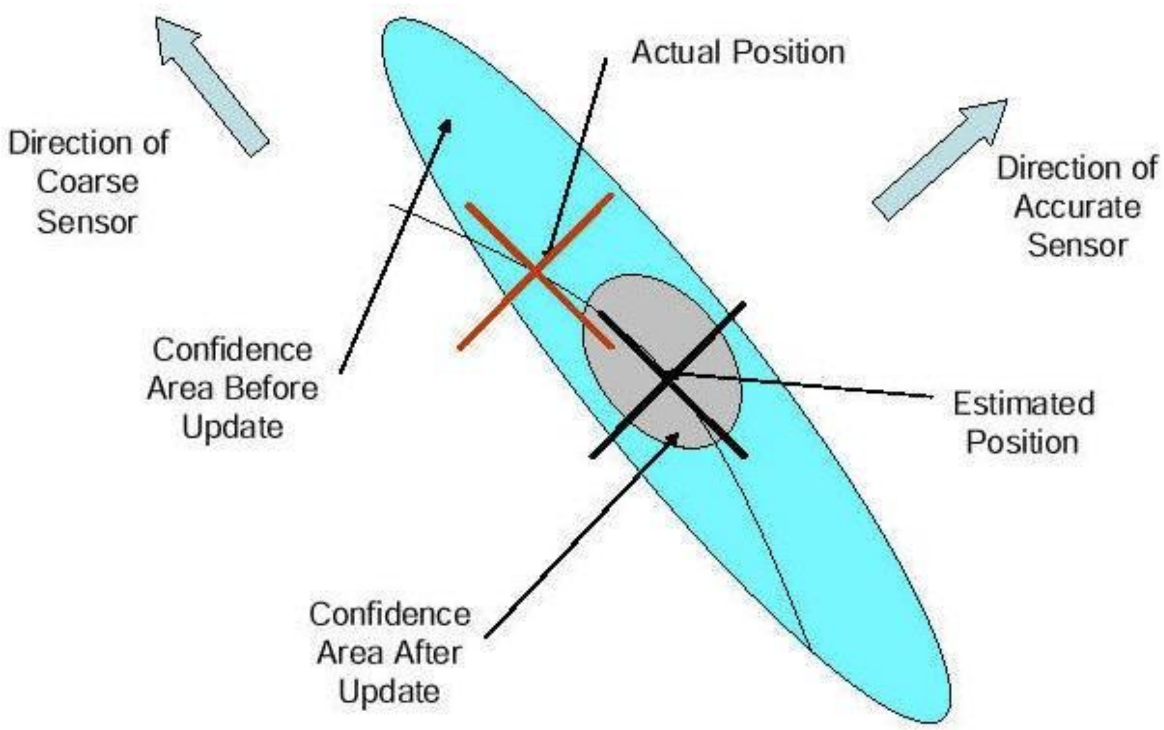
- Divergence results because of an over-reduction of the state covariance matrix which is particularly pronounced in accurate measurement directions.

  -hence, the true state vector lies outside the ellipsoid that approximates the confidence area of the estimated state vector in just a few iterations.

The Mechanism of Divergence



The Divergence Mechanism (cont'd)



'Standard' Remedies

- Various 'Bump-Up' Strategies for the EKF that artificially increase the covariance (Perea, How, Breger, Elosegui, AIAA Conference, 2007)

- Variations of the GSF(especially the modified truncated GSF)

- Various 'Bump-Up' Strategies for the UKF (e.g. B-UKF)

Bottom Line: Remedies attempt to 'ad hocly' ensure a reasonable size covariance by 'bumping it up'.

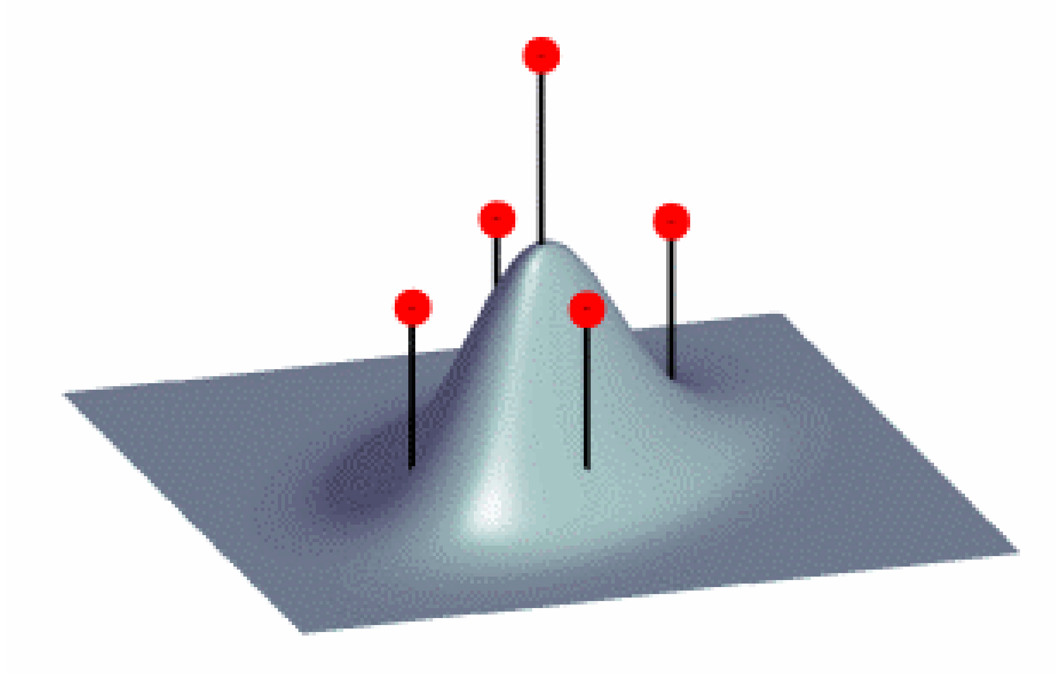
The Principle of the UKF

- It approximates, instead of the nonlinear function, its distribution.

- In contradistinction to particle filters, which approximate the full probability distribution, the UKF avoids such complexity by generating a set of ‘well distributed’ points around current estimate , propagating this set of ‘sigma’ points to the next epoch, and re-computing the statistics from the resulting set of ‘sigma’ points.

The General Principle of the UKF

- The trick is how to pick the ‘sigma’ points so as to most accurately represent the desired probability distribution



The Unscented Transformations

- Represent various ways of ‘picking’ a finite number of ‘sigma’ points (particles) :

  -Standard Unscented Transformation

  -Generalized Unscented Transformation

  - Scaled Unscented Transformation

Bottom Line: they all attempt to ‘dilate’ the confidence area in a min-max way so as to:

-include the true mean value of estimate

-avoid skewed overcorrection by maintaining a reasonable ’measure’ of confidence area

The Standard Unscented Transformation

- 2n sigma points are chosen such as:

$$X_i = \bar{x} + (\sqrt{nP})_i$$

$$X_{i+n} = \bar{x} - (\sqrt{nP})_i$$

These sigma points are then transformed via propagation dynamics, and the measurement function, with update ensuing

$$X_i^- = f(X_i, t) \qquad i = 1,2 \ldots .2n$$

$$Y_i = g(X_i) \qquad i = 1,2 \ldots .2n$$

The Generalized Unscented Transform

- 2n+1 sigma points are chosen with weights w_i

$$X_0 = \bar{x} \qquad W_0 = \frac{k}{n+k}$$

$$X_i = \bar{x} + (\sqrt{nP})_i \qquad W_i = \frac{1}{2(n+k)} \qquad i = 1,..n$$

$$X_{i+n} = \bar{x} - (\sqrt{nP})_i \qquad W_{i+n} = \frac{1}{2(n+k)}$$

The Scaled Unscented Transform

- The choice of ‘sigma’ points is :

$$X_0' = X_0$$

$$X_i' = X_0 + a(X_i - X_0)$$

$$X_{i+n}' = X_0 + a(X_{i+n} - X_0)$$

$$W_0' = \frac{W_0}{\alpha^2} + (1 - \frac{1}{\alpha^2})$$

$$W_i' = W_i / a^2 \qquad i = 1, .. n$$

$$W_{i+n}' = W_{i+n} / a^2$$

Considerations and Comments

- Choice of ‘right’ values for the regulating parameters k,α, requires a large number of experimentations/calibrations with each new application:

    -trial and error approach; no guarantees

    -not suitable for real-time applications

    -in addition, need to ‘view’ nonlinearity from a distance, so as to better reflect its nature in the update, which affects covariance

The New Algorithm

- Is based on a SVD of P, which, given its symmetry and positive semi-definiteness, can be simply realized
- It is numerically robust
- Results in sigma points exactly along the basic axes of the confidence area, hence also following closely its evolution, thus covering its vital space with no redundancy but with completeness

The New Algorithm

- With

$$P^{SVD} = WSV^T \Rightarrow P^{SVD} = \sum_{i=1}^{n} s_i\, w_i v_i^T$$

The generalized 'sigma' points are chosen as:

$$X_0 = \bar{x} \qquad W_0 = \frac{k}{(n+k)}$$

$$X_i = \bar{x} + \sqrt{(n+k)\lambda_\iota}\,.u_i \qquad W_i = \frac{1}{2(n+k)} \qquad i = 1,..n$$

$$X_{i+n} = \bar{x} - \sqrt{(n+k)\lambda_\iota}\,.u_i \qquad W_{i+n} = \frac{1}{2(n+k)}$$

Sigma Points Choice Comparison

- Transitioning from polar measurements to cartesian state space:

$$(r, \theta) \xrightarrow{h} (x, y)$$

We have:

$$r \sim N(\bar{r}, \sigma_r) \qquad \bar{r} = 1m \qquad \sigma_r = 0.01m$$

$$\theta \sim N(\bar{\theta}, \sigma_\theta) \qquad \theta = 90^0 \qquad \sigma_\theta = 20^0$$

Where there are orders of magnitude discrepancy in sensor accuracy

Example for ‘Sigma’ Points Choice

- For the polar coordinates the ‘sigma’ points are chosen as:

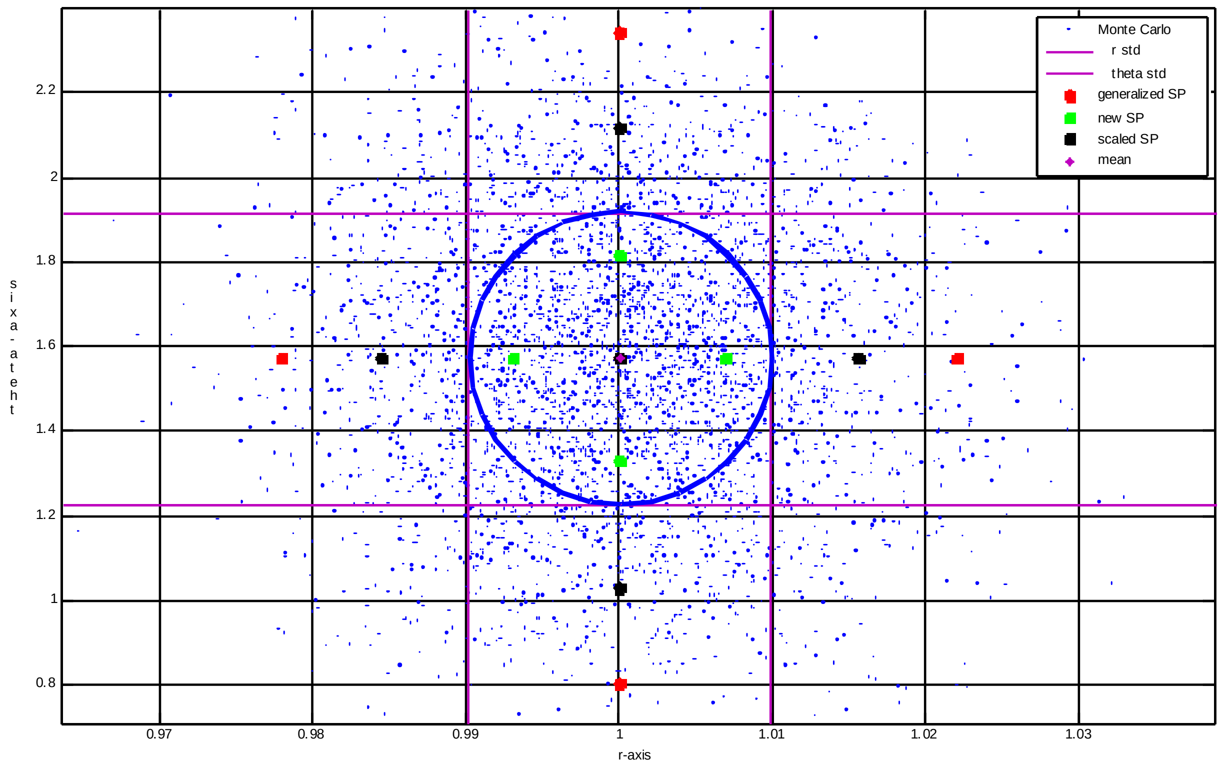
$$\begin{pmatrix} r_1^c \\ \theta_1^c \end{pmatrix} = \begin{pmatrix} \bar{r} \\ \bar{\theta} \end{pmatrix} + \left(\frac{\sigma_r}{\sqrt{2}}\right)\begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} r_2^c \\ \theta_2^c \end{pmatrix} = \begin{pmatrix} \bar{r} \\ \bar{\theta} \end{pmatrix} + \left(\frac{\sigma_\theta}{\sqrt{2}}\right)\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} r_3^c \\ \theta_3^c \end{pmatrix} = \begin{pmatrix} \bar{r} \\ \bar{\theta} \end{pmatrix} - \left(\frac{\sigma_r}{\sqrt{2}}\right)\begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} r_4^c \\ \theta_4^c \end{pmatrix} = \begin{pmatrix} \bar{r} \\ \bar{\theta} \end{pmatrix} - \left(\frac{\sigma_\theta}{\sqrt{2}}\right)\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

HSS

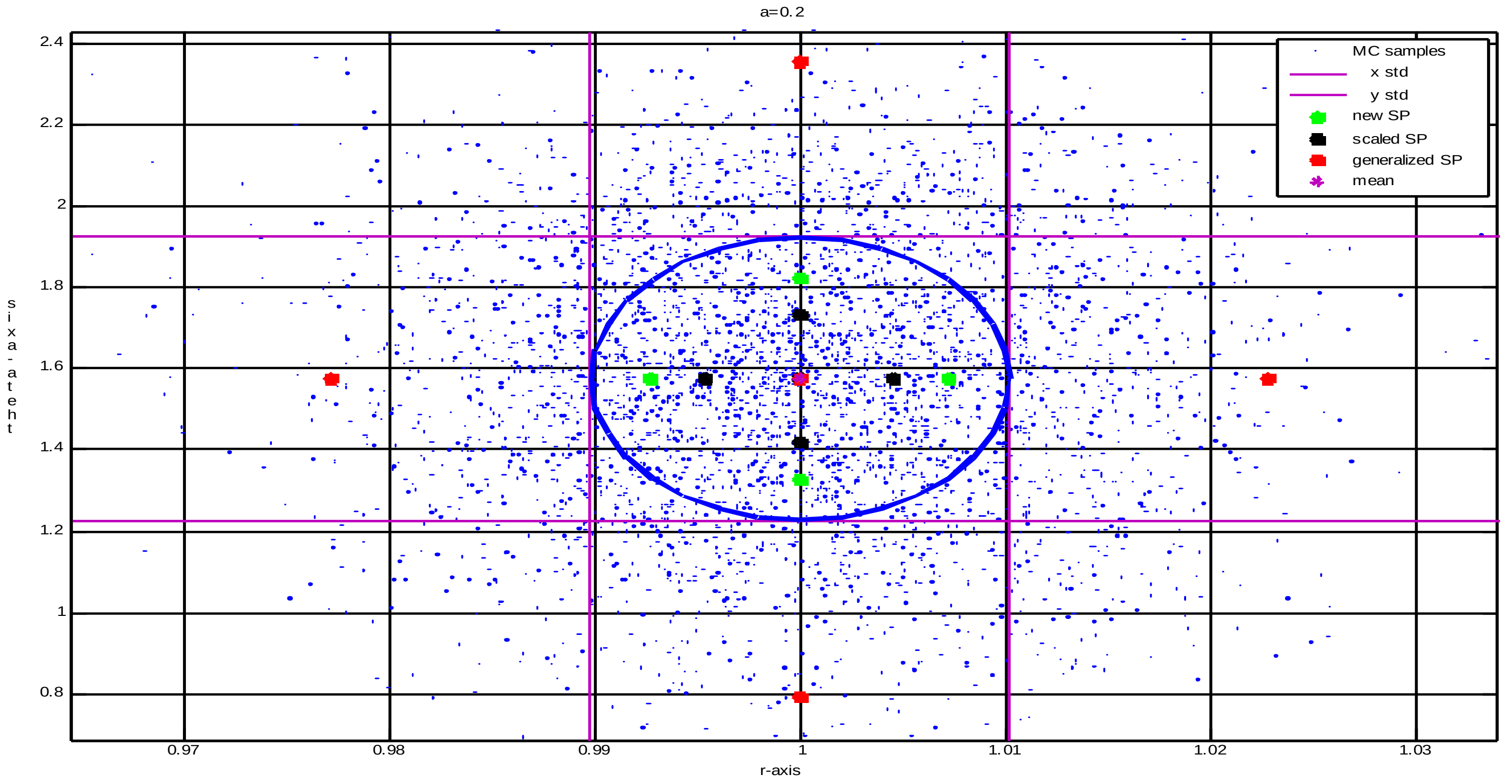
Comparison of Sigma Point Positions

Measurement Space(r, angle θ)with α=1/sqrt(2)



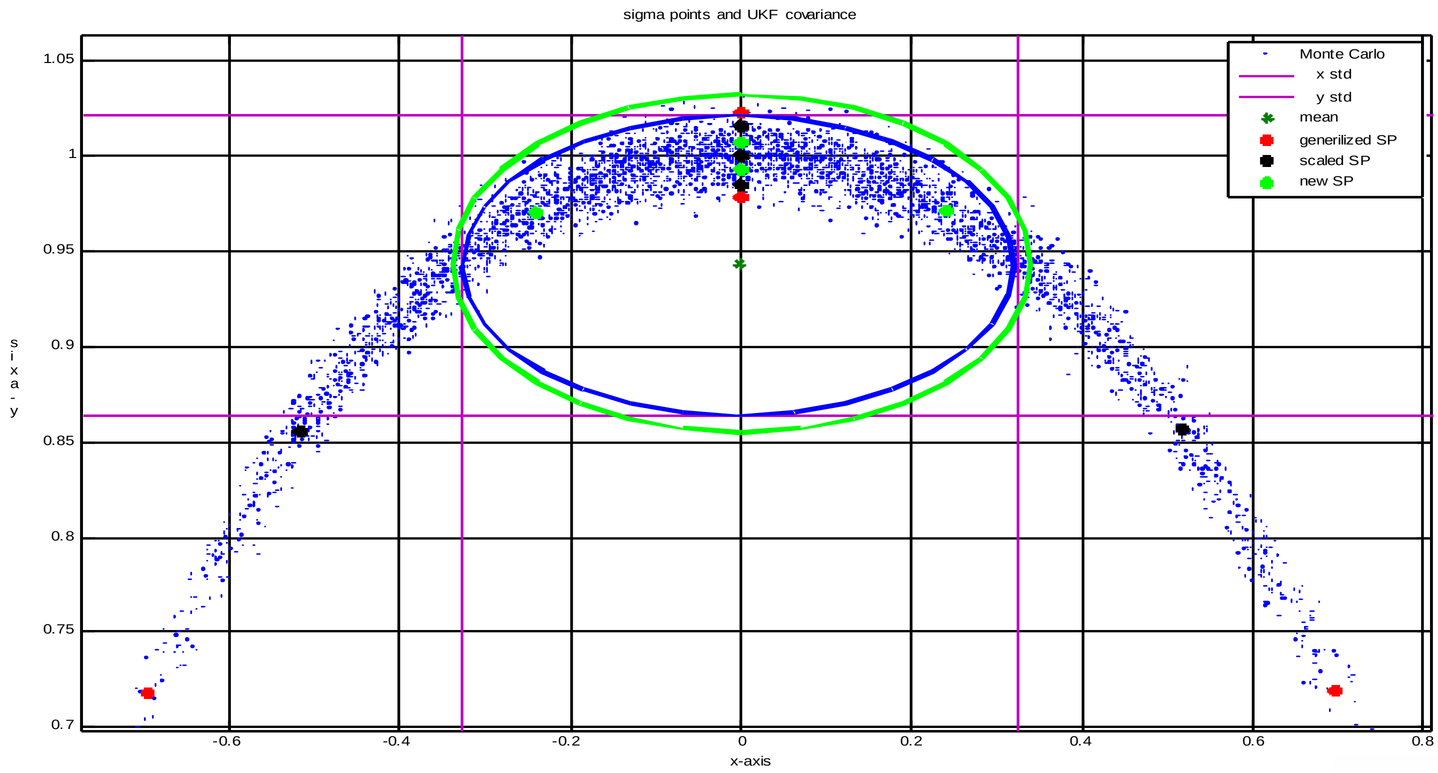
Comparison of ‘Sigma’ Point Positions

- Measurement Space (radius , angle) with α=.2



Comparison of ‘Sigma’ Point Positions

State Space (x , y) with α = 1/sqrt(2)



Challenges

- **Topology / Architecture**
  - how is the interconnection (of 'modules') impacting the decision/control process

- Tradeoffs how **decision quality** influences **information quantity**

- **Value of Information**
  -understanding and characterizing the information that can best support a decision/control or learning objective

Challenges (cont'd)

- Decisions/Control under limited information
  - -what are the fundamental limits
  - -what are the impacts of partial information

- **Complexity** is a significant issue : it inextricably links many disciplines
- **Synergy between** computation, information/ communi-cations and control is an absolute must

Conclusions

- The range of challenges/applications is vast and the technology enablers limitless but systematic synthesis tools lacking

- **Complexity** is a significant issue : it inextricably links many disciplines
- **Synergy** between computation, information/ communi-cations and control is an absolute must