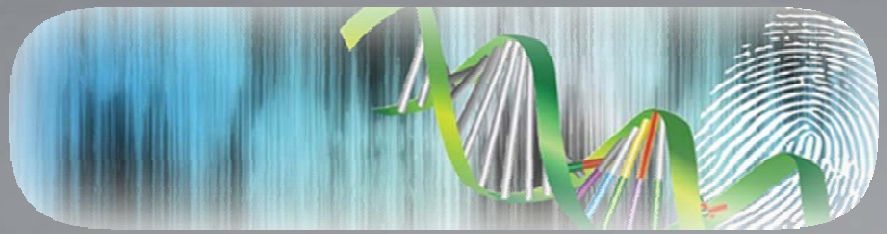
Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

PNNL-SA-102409

A Systematic Process for Discovering New Signatures

MARK F. TARDIFF
SIGNATURE DISCOVERY INITIATIVE

Pacific Northwest National Laboratory
ADSA10, Boston
May 6, 2014



The SDI Objective

- Develop a formal process for signature discovery
- Transform multi-INT signature development to become
  - ***More efficient*** – by reducing trial-and-error and decreasing the time to discovery a signature
  - ***More economical*** – by delivering methods and tools that allow users to reuse, rather than reinvent, the resources needed to construct, detect, and validate signatures
  - ***More rigorous*** – by providing robust and well-defined processes for signature discovery

The Cargo Screening Dilemma

- Many targets with complementary or competing approaches
  - X-Ray technologies
  - Magnetics
  - Heartbeat monitoring
  - K-9 for explosives, humans
  - CO2 monitoring
  - Ion mobility spectrometry
  - Millimeter wave imaging
  - Terahertz methods
  - Emergent passive and active interrogation techniques

The Cargo Screening Dilemma
Continued

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Operational constraints
  - Commerce is King
  - Time to screen
  - Time to confirm
  - Flight schedules
  - No control of when cargo arrives at the warehouse

- All in the context of extremely rare true positives – predisposed to believe that a detect is a false positive
- Containing costs is paramount because of prevalence – don't spend a lot of money on what doesn't happen

The Cargo Screening Dilemma
Continued

Boudhanath Stupa
Kathmandu, Nepal

The Cargo Screening Dilemma
Continued

A Signature Centric Approach

- To a hardware developer the cargo screening challenge is a hardware problem
- To a software developer the challenge is a software problem

- From a signature discovery perspective, cargo screening is a classifier problem

Introduction to Signature Discovery

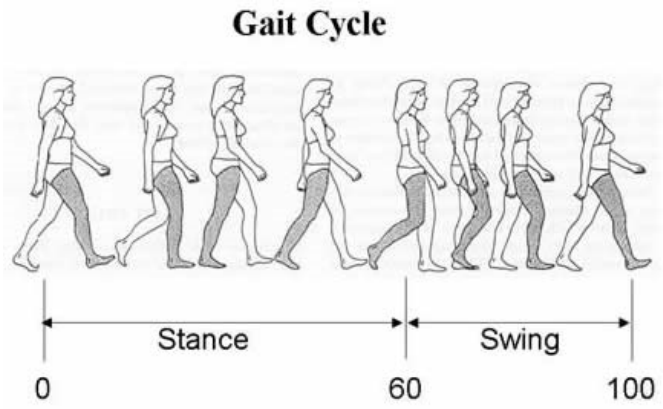
- ***Signature***: A distinguishing collection of features that *characterizes* a phenomenon of interest
- Forensic, diagnostic, prognostic
- ***Example***: gait analysis for biometrics
  - Static measurements
  - Dynamic measurements
  - Other measurements?

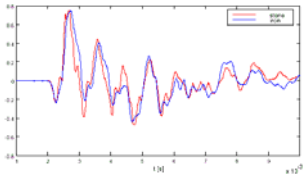
Feature	Measure-ment	Prob of ID
Step length	24 ± 2 in	
Step width	4.0 ± 0.5 in	
Knee angle	142 ± 5 deg	60%
Walking speed	5.2 ± 0.2 fps	
Cycle time	2.6 ± 0.1 s	80%
Acoustic power	70 dB	95%

Is this Kim Smith?

Construction	Validation
How were these features selected?	Is this a "good" signature?
Dynamics	**Detection**
What's the "shelf life" of this signature?	Can we find this signature in a crowd?

?

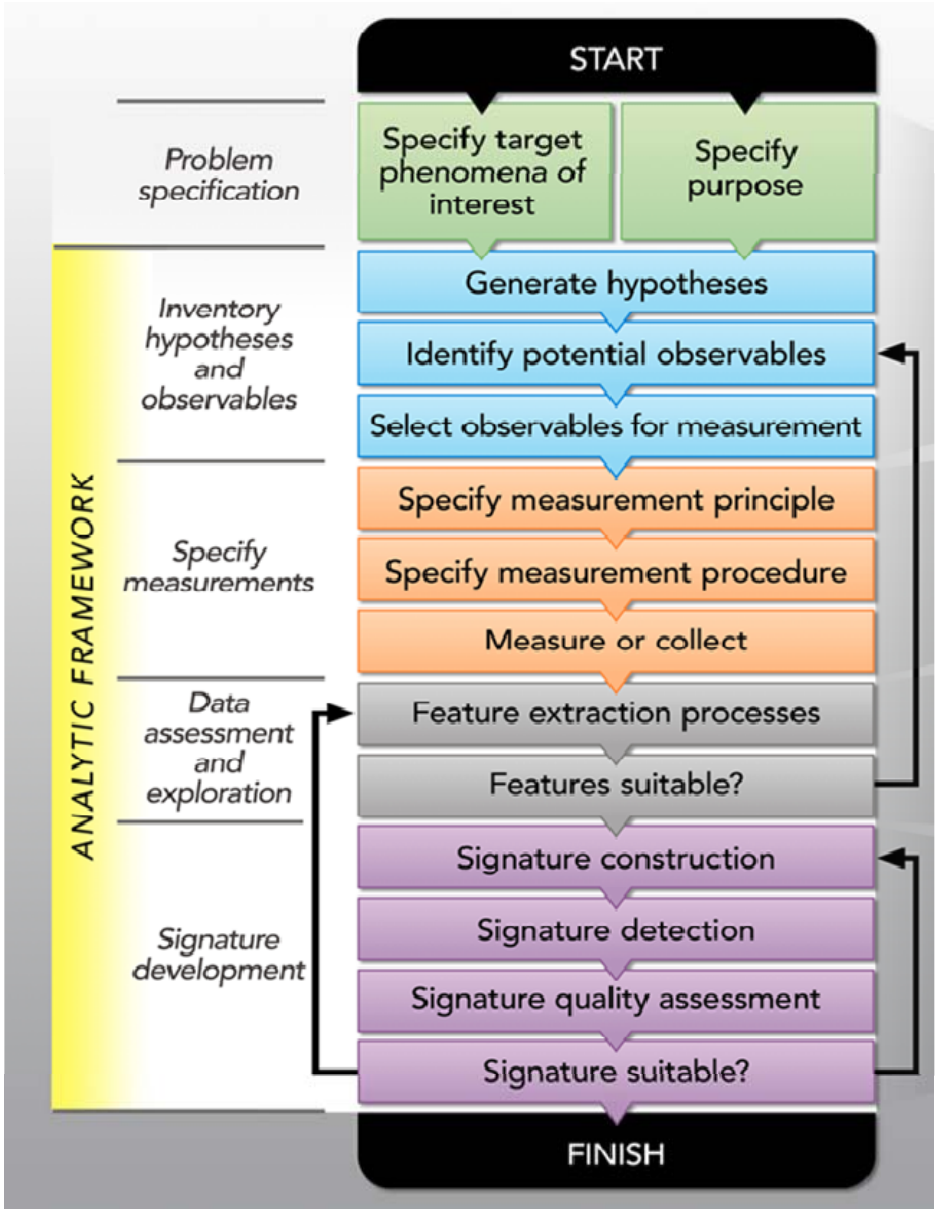


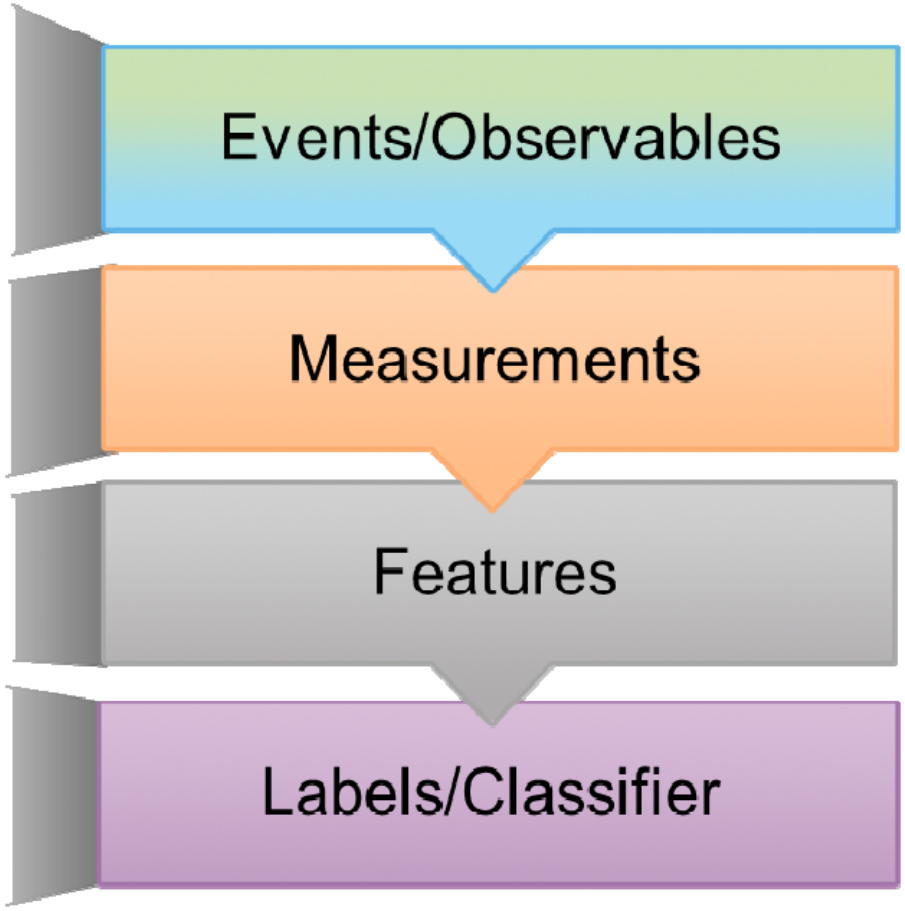


Open Questions in Signature Discovery

- Challenges
  - Modern signatures often involve features from many different data types
  - Signatures are most useful when they are easily interpreted by decision-makers
- Key questions
  - How do we best select features from multiple measurement sources to construct our signature?
  - How do we detect signatures in “real world“ environments?
  - How do we assess the quality of a signature and compare different signatures?
  - How do we recognize change and adapt signatures to dynamic phenomena?
  - *Is this process generalizable across domains?*

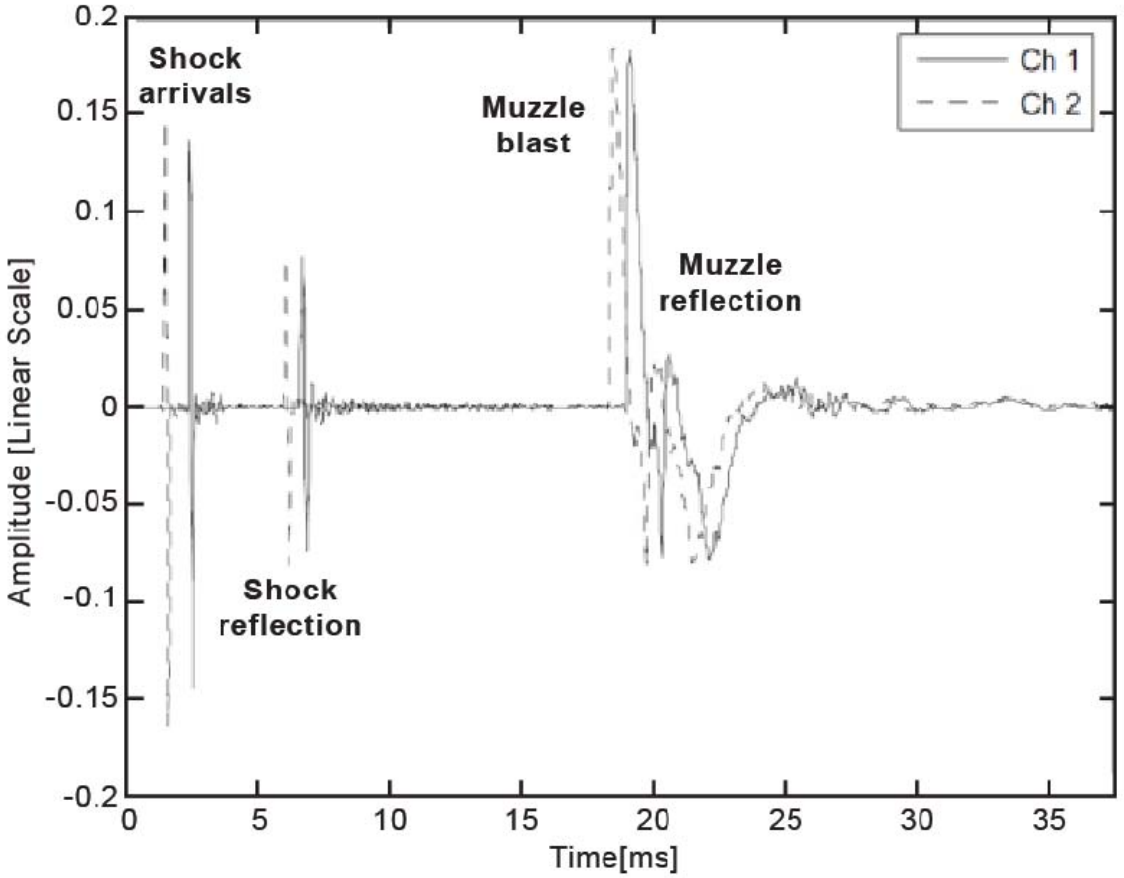
Signature System





Methodology
Gunshot Example

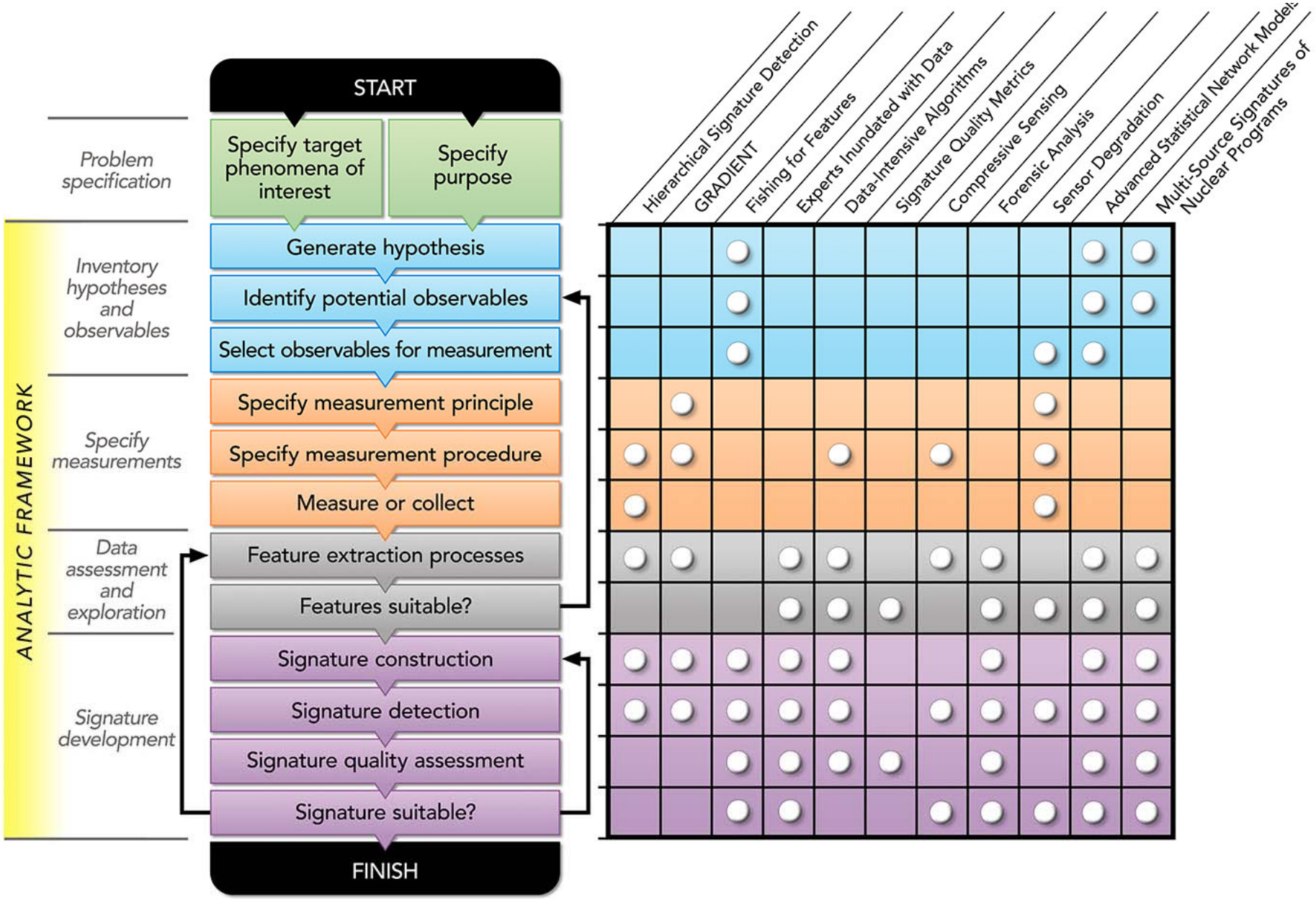
- Events/observables:
  - Gunshots
- Data
  - Sound collected over the dynamic range of the microphone
- Features
  - Frequency intervals that are related to gunshots, time intervals for gunshot duration
- Classifier
  - {Gunshot, no gunshot}



Events $\xrightarrow{\mu}$ Measurements $\xrightarrow{\eta}$ Features $\xrightarrow{\delta}$ Labels

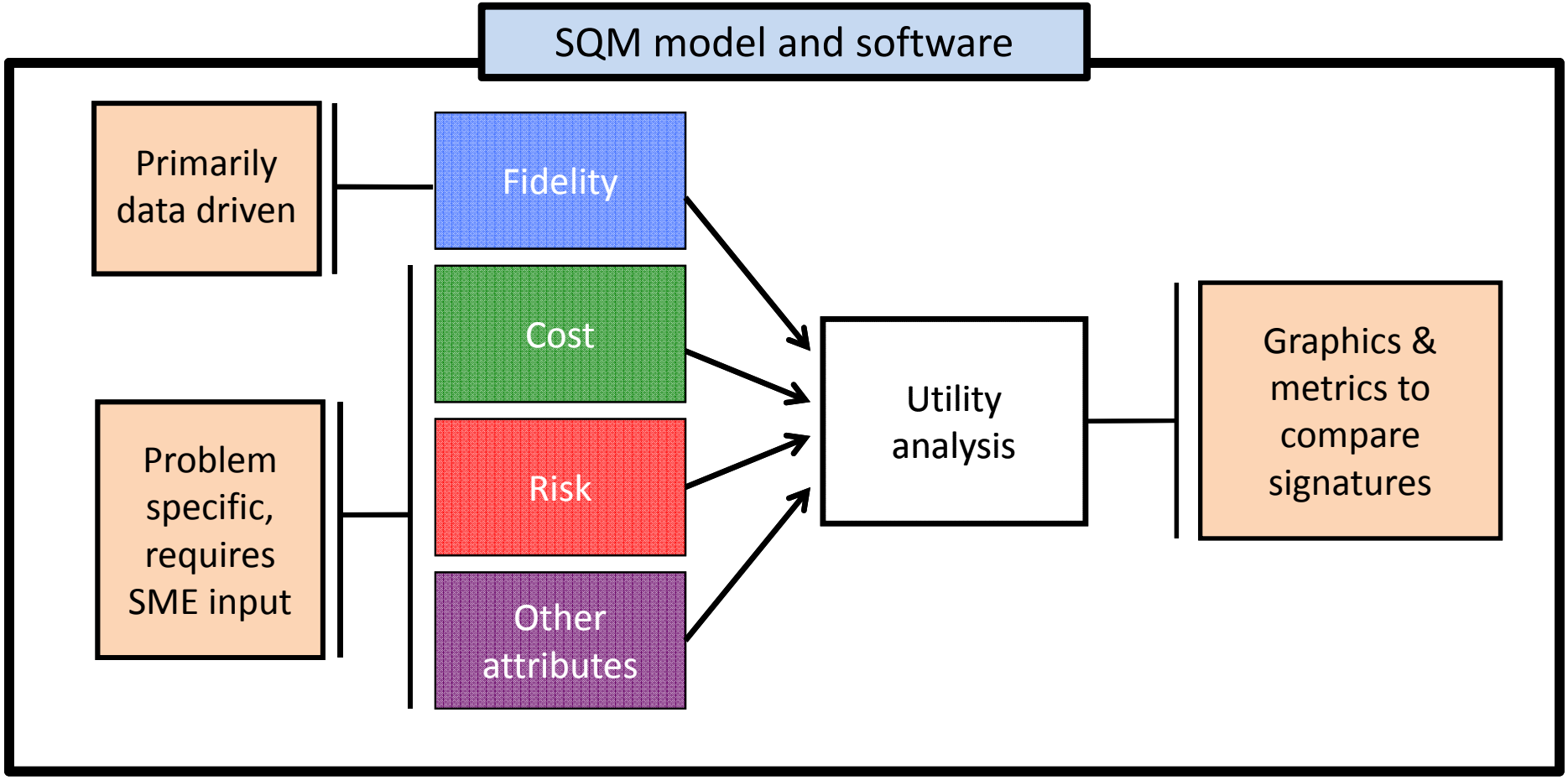
Generalized Form	Where
$\mu: E \rightarrow M$ $\eta: M \rightarrow F$ $\delta: F \rightarrow L$	E = Events, $\{e_1, e_2, e_3 \ldots\}$ M = Measurements, $\{m_1, m_2, m_3 \ldots\}$ F = Features, $\{f_1, f_2, f_3 \ldots\}$ L = Labels, $\{l_1, l_2, l_3 \ldots\}$

Project Portfolio

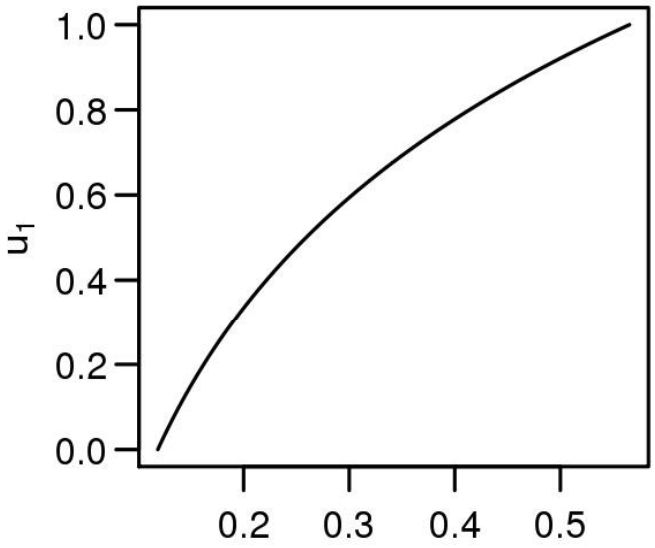


Signature Quality Metrics
Conceptual model

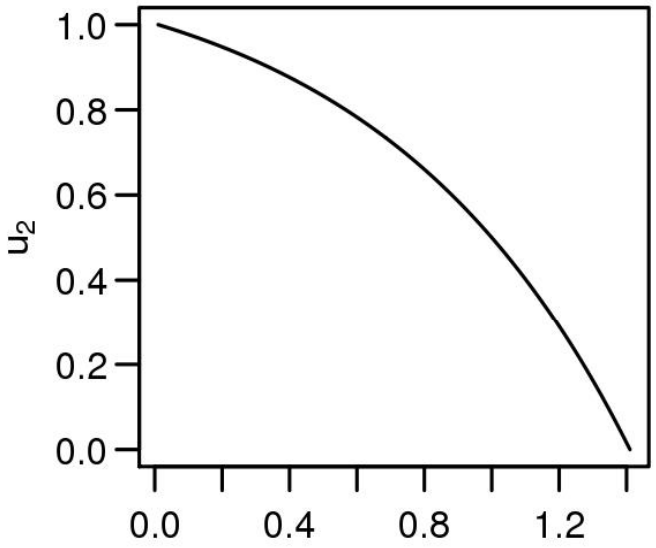
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965



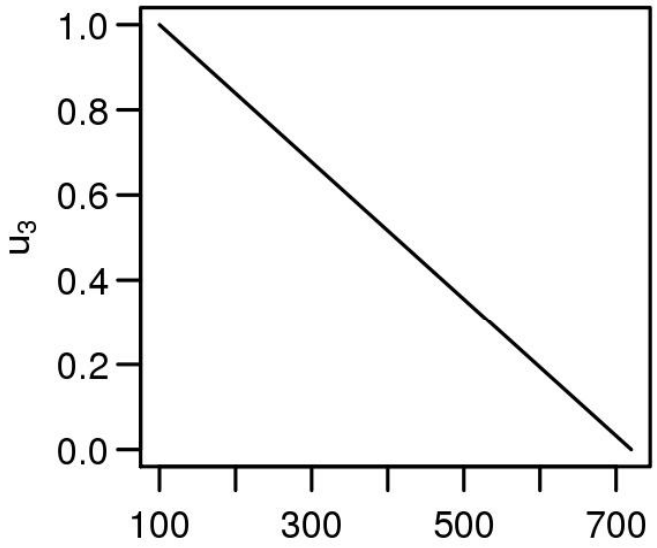
Calculating Utility



Attribute 1 - Probability of correct outcome



Attribute 2 – Dwell Time



Attribute 3 – Training cost

Additive utility function:

Utility = $0.85\ u_1$ (Probability of correct outcome) +
$0.15\ u_2$ (Dwell time) +
$0.05\ u_3$ (Training cost)

$$u(\mathbf{z}_{ij}, \theta_j, S_i(\mathbf{x}_j)) = \alpha_1 u_1(\theta_j, S_i(\mathbf{x}_j)) + \sum_{a=2}^{A} \alpha_a u_a(z_{a,ij})$$

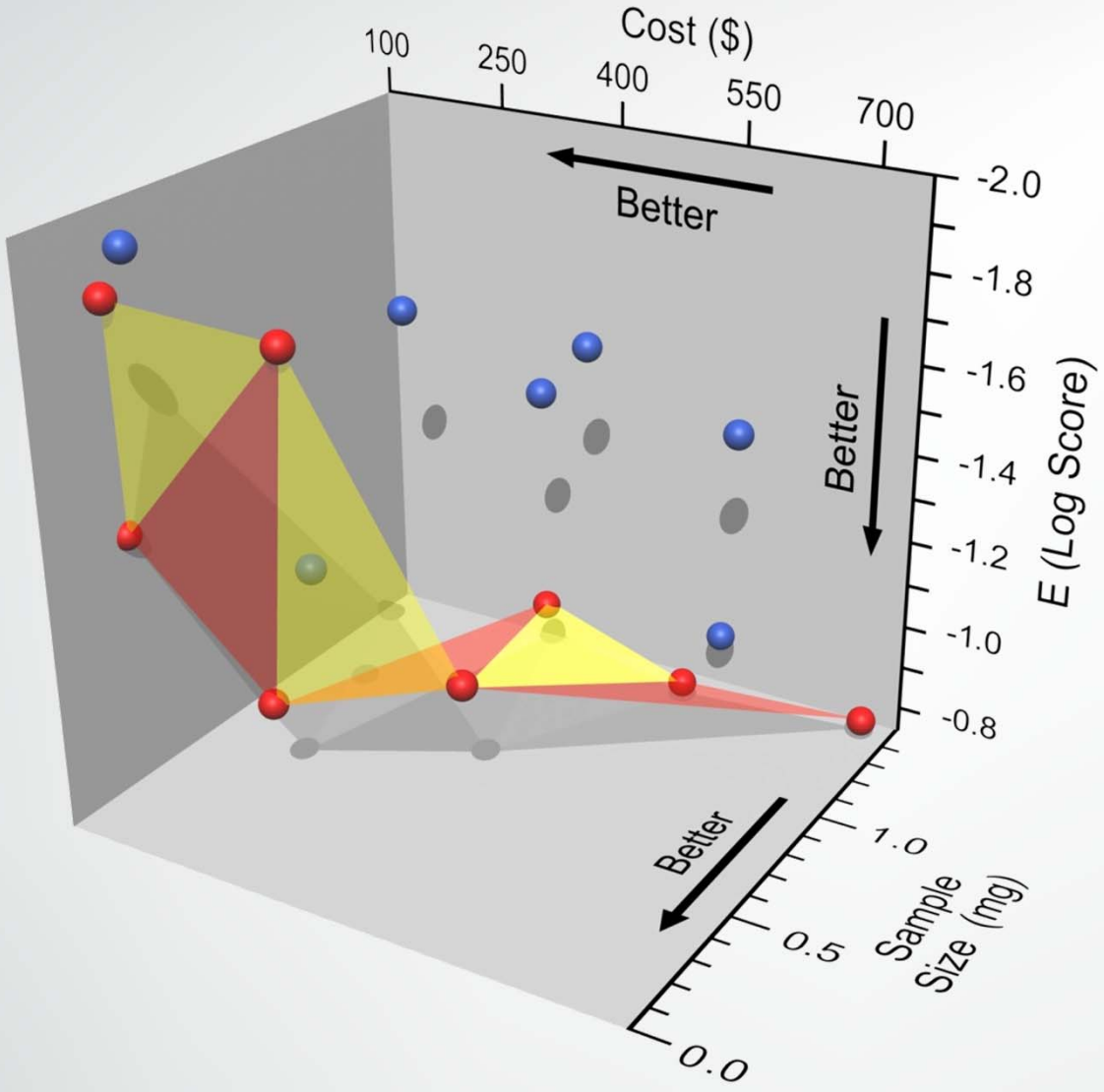
Calculating Expected Utility

- Expected Utility, $E(u(S))$, can be estimated by
  - Calculating the additive utility for each observation in the test data
  - Calculating the weighted average of the resulting utilities (weighted by the likelihood of each observation)

$$\vartheta(S_i) = \iiint u(\mathbf{z}_i, \theta, S_i(\mathbf{x}))\ dF_{\mathbf{Z}|\mathbf{X}\theta} dF_{\mathbf{X}|\theta} dF_{\theta}$$

- System S_1 is preferred to S_2 if $E(u(S_1)) > E(u(S_2))$

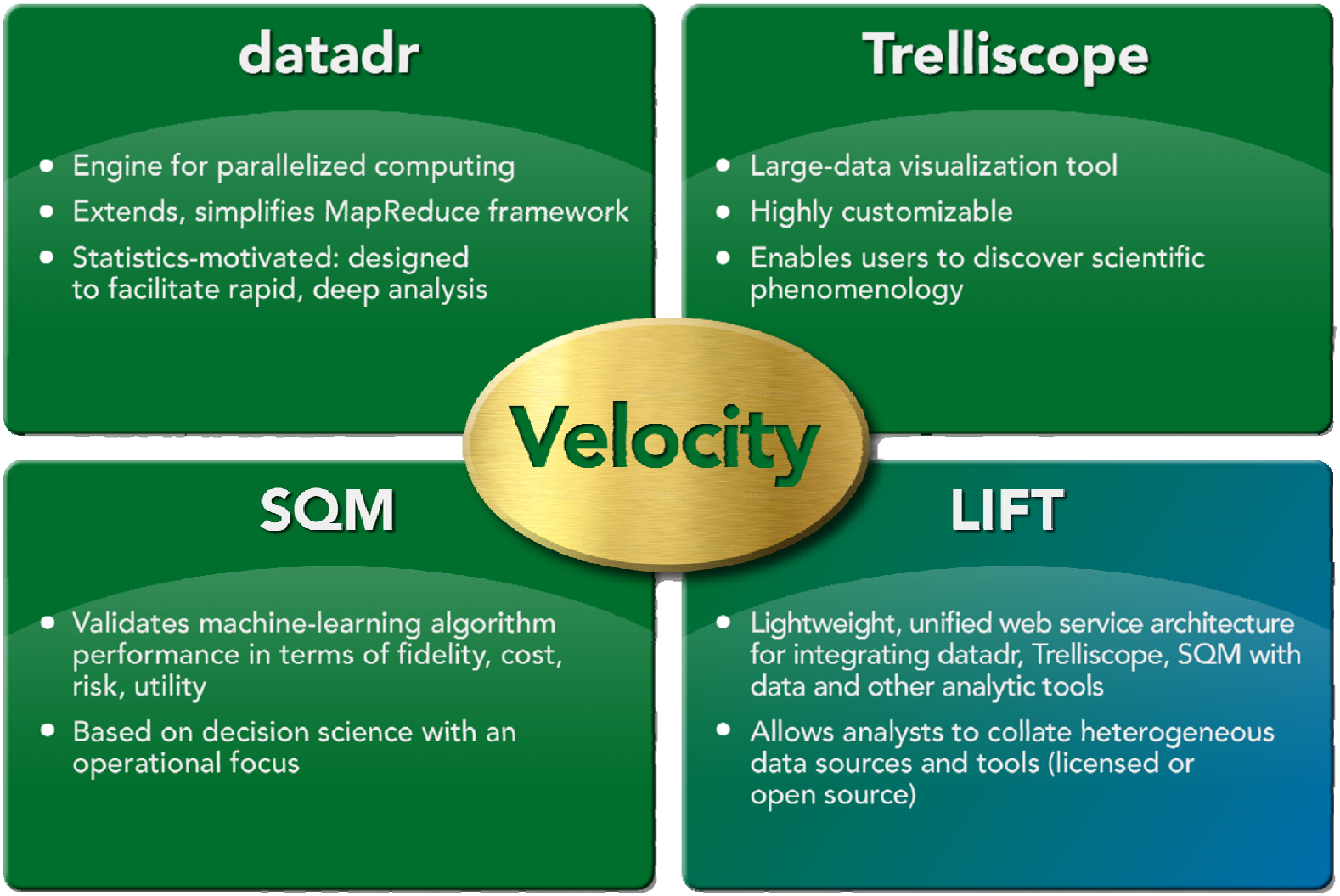
Pareto Frontier



Additional Signature Discovery Capabilities

- Which variables in my data relate best to my target?
- Under what conditions does sensor performance fall off?
- Can we sample data as it's collected and reconstruct the original?
- How do we influence machine learning with expert opinion?
- Can we initiate data processing at the sensor and reduce I/O demands?
- Can we merge measurement and unstructured data and assign confidence to outcomes?
- How do we assess relative similarity in changing environments?
- Are there ways to graph complex data and detect changes over time?
- Can we construct a tools environment for signature discovery?
- Can we ease user access to signature discovery tools?

Velocity



DHS, HSARPA – Steve Dennis and Meredith Lee

Questions

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Mark F. Tardiff
Pacific Northwest National Laboratory
509-375-2530
mark.tardiff@pnnl.gov