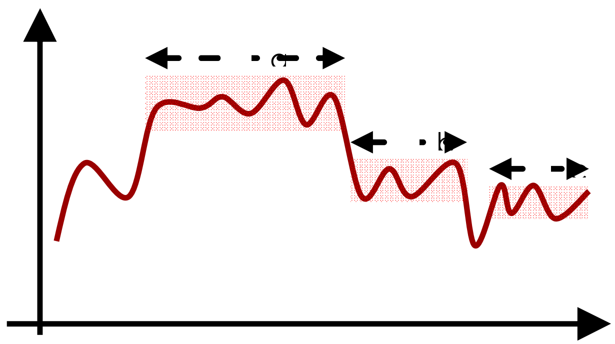
Basic problem

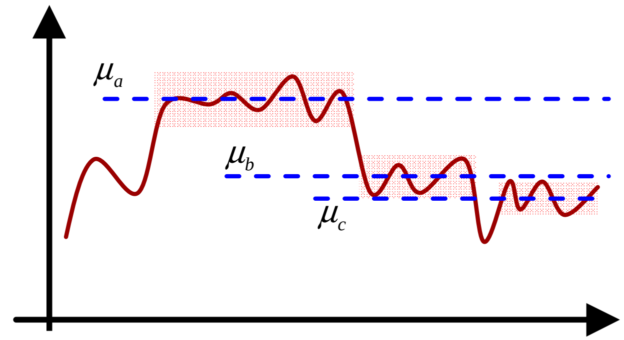
- Want to detect approximate bursts of approximate stability
  - Burst is of unknown duration
  - Burst has unknown mean and sigma
  - Distribution of the original signal may be unknown
  - Burst may contain outliers with respect to samples in the burst
  - Burst may be part of a larger burst

- Want model and device that provides measure of confidence
  - Over the error of the imposed approximated model over the derivative form of the signal

Measurements Setup



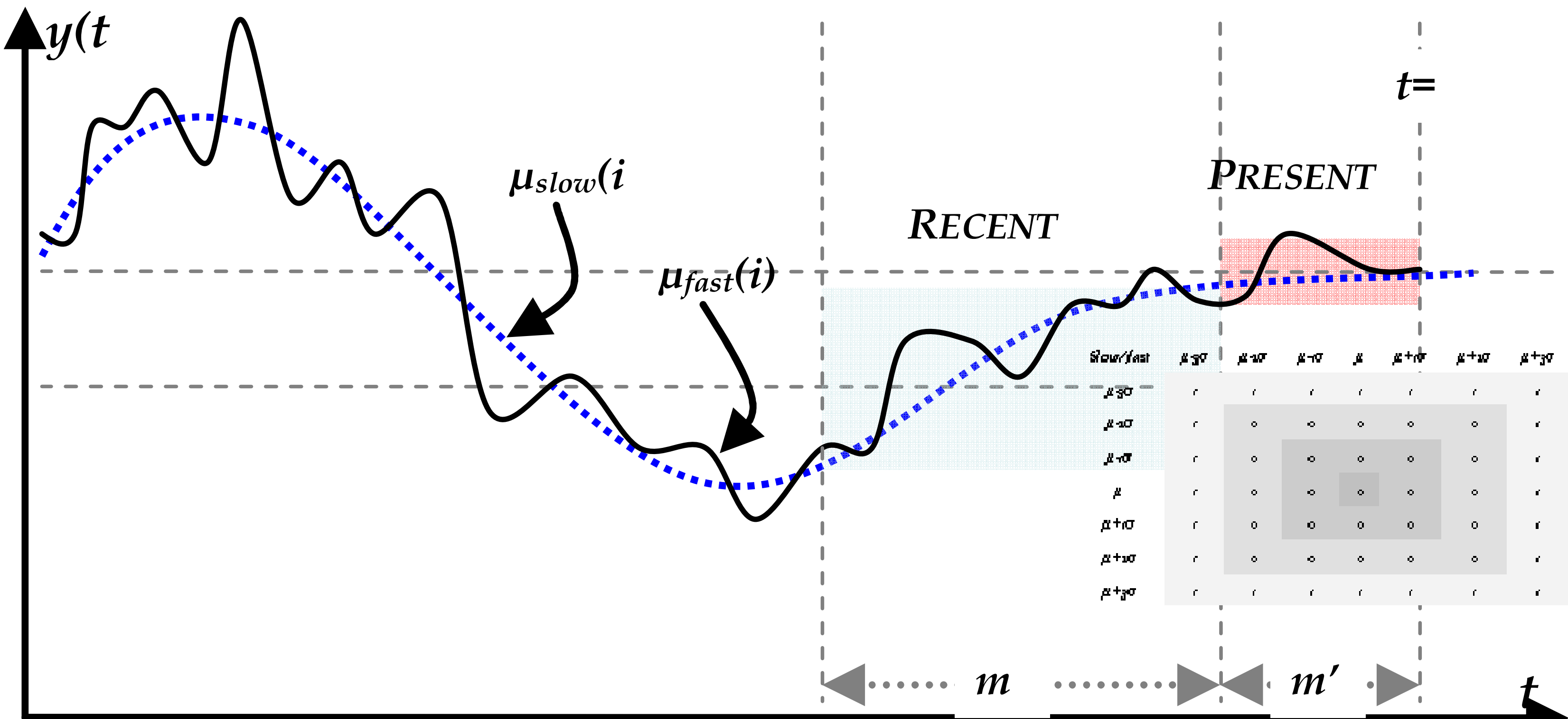
(a)



(b)

- Want to
  - estimate at time *t*, presence of a possibly underlying stationary burst
  - estimate at time *t*, targeting mean of such underlying burst
  - have some confidence that stationary burst model fits data interval
  - accelerated detection of departure from stationary burst model to detect misalignment to the hypothesis
  - operating region behavior to parameterize quality/quantity of bursts

Decision Making Problem: Similarity of PAST to PRESENT



View into the setup of hypothesis testing for "approximate τ-invariance" at time index *i.*

Dr. Nelson R. Manohar-Alers

Approach

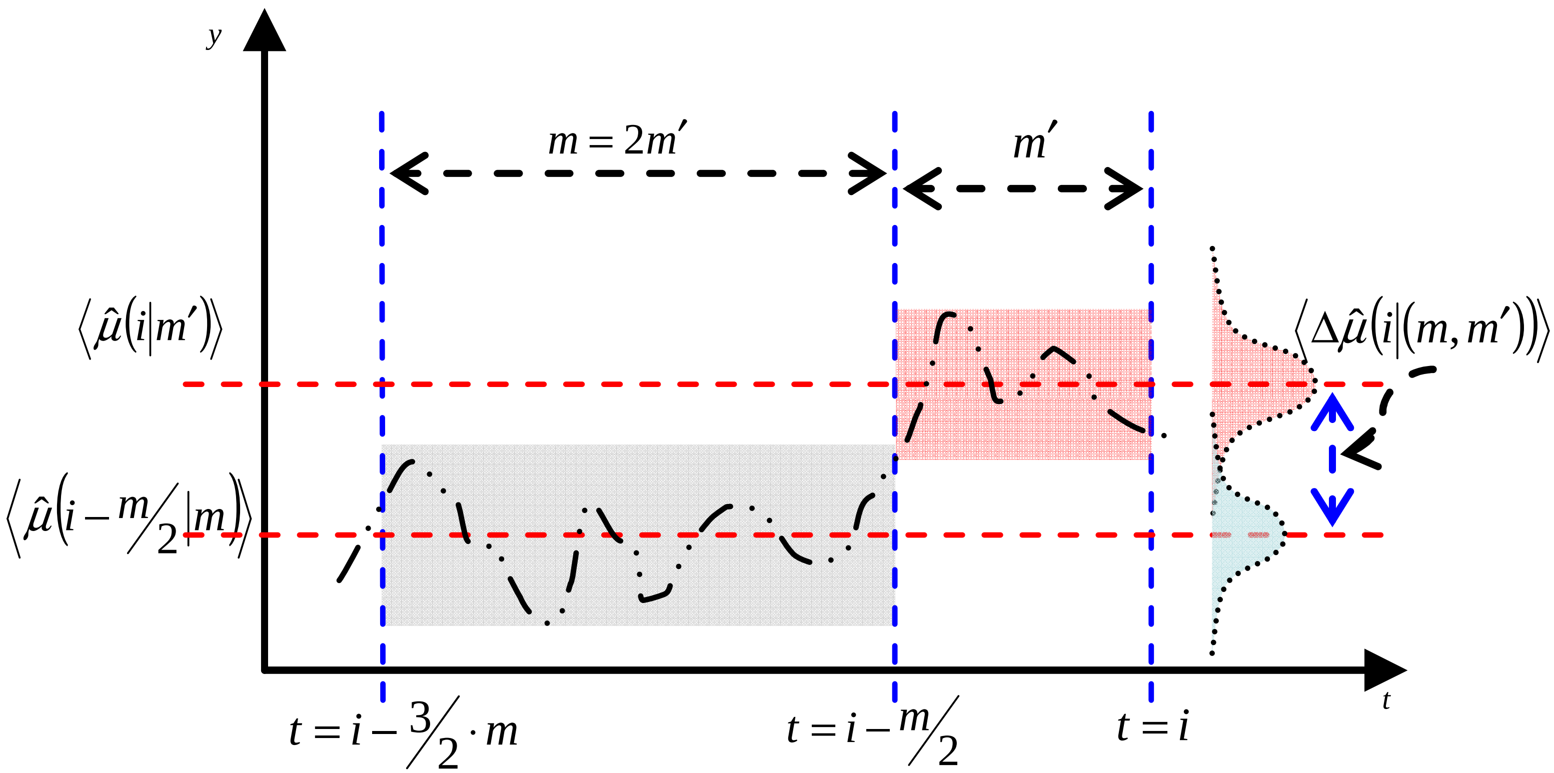
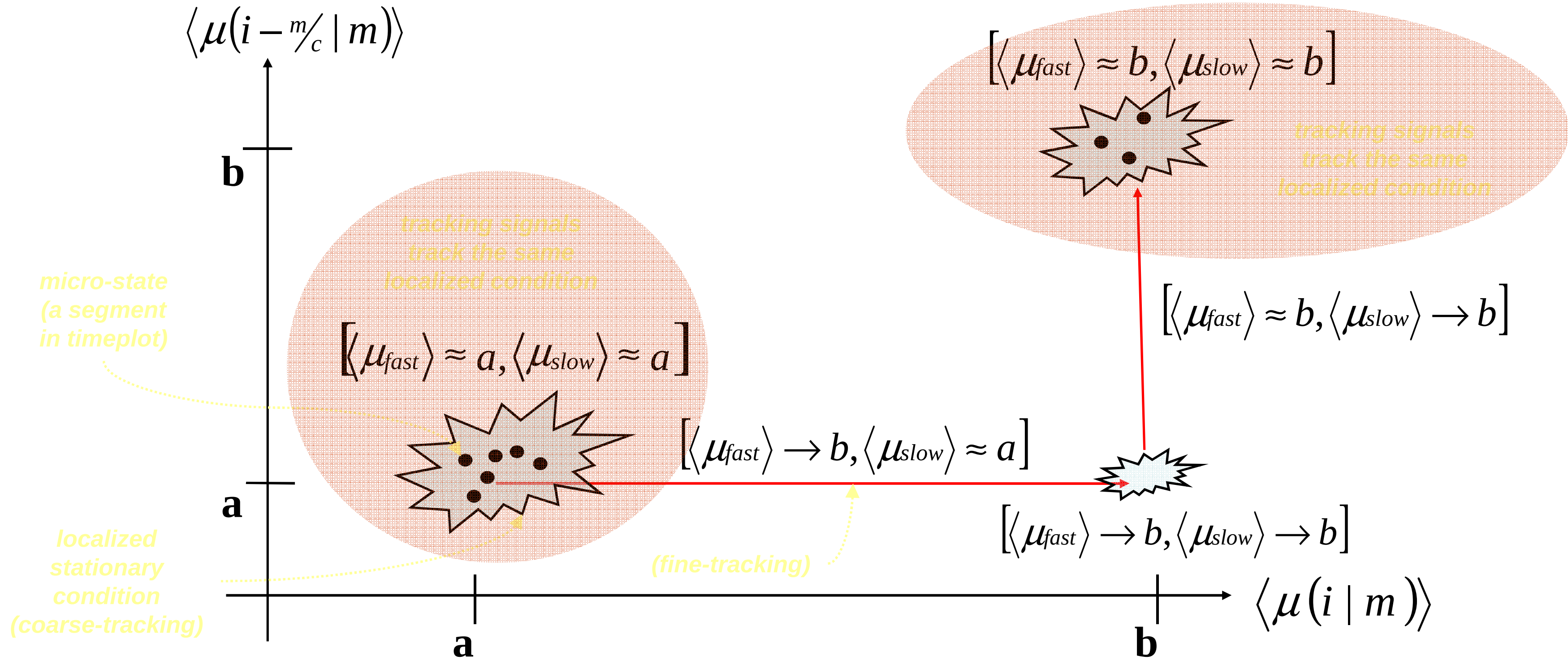


Fig. 8: An inferential approximation to *SIMILAR()*. Testing setup for the HPS conjecture at some time *i w.r.t.* the outlooks of the CLT-stabilized signals.

Space spanned by PAST and PRESENT signals



- When BOTH outlooks WITHIN <u>localized stationary condition</u>
  - inside a tightly knit coarse-tracking cluster
  - samples true mean of localized stationary condition
- once ANY of the outlooks LEAVES localized stationary condition
  - manifests as fine-tracking transition away from coarse-tracking cluster

MSE Equivalency Result (Approximate Presence of Approximate Stationary)

- Let $\langle g(i) \rangle$ be an arbitrary signal,
- Let $\langle g(i-\tau) \rangle$ be its $\langle \tau \rangle$-delayed version.
- Let $\langle f(i-\tau) \rangle$ be a CLT-smoothed version of a $\langle \tau \rangle$-delayed $\langle g(i-\tau) \rangle$.
- Let α be a confidence level.
- Let $\langle v \rangle$ be an arbitrary finite interval of size m'.

- Then, at an α confidence level, the maximum error permissible $\langle MSE_{max}(i) \mid m' \rangle$ along an interval $\langle v \rangle$ of signal $\langle f(i) \rangle$ if **approximate τ-invariance** exists across interval $\langle v \rangle$ of signal $\langle f(i) \rangle$

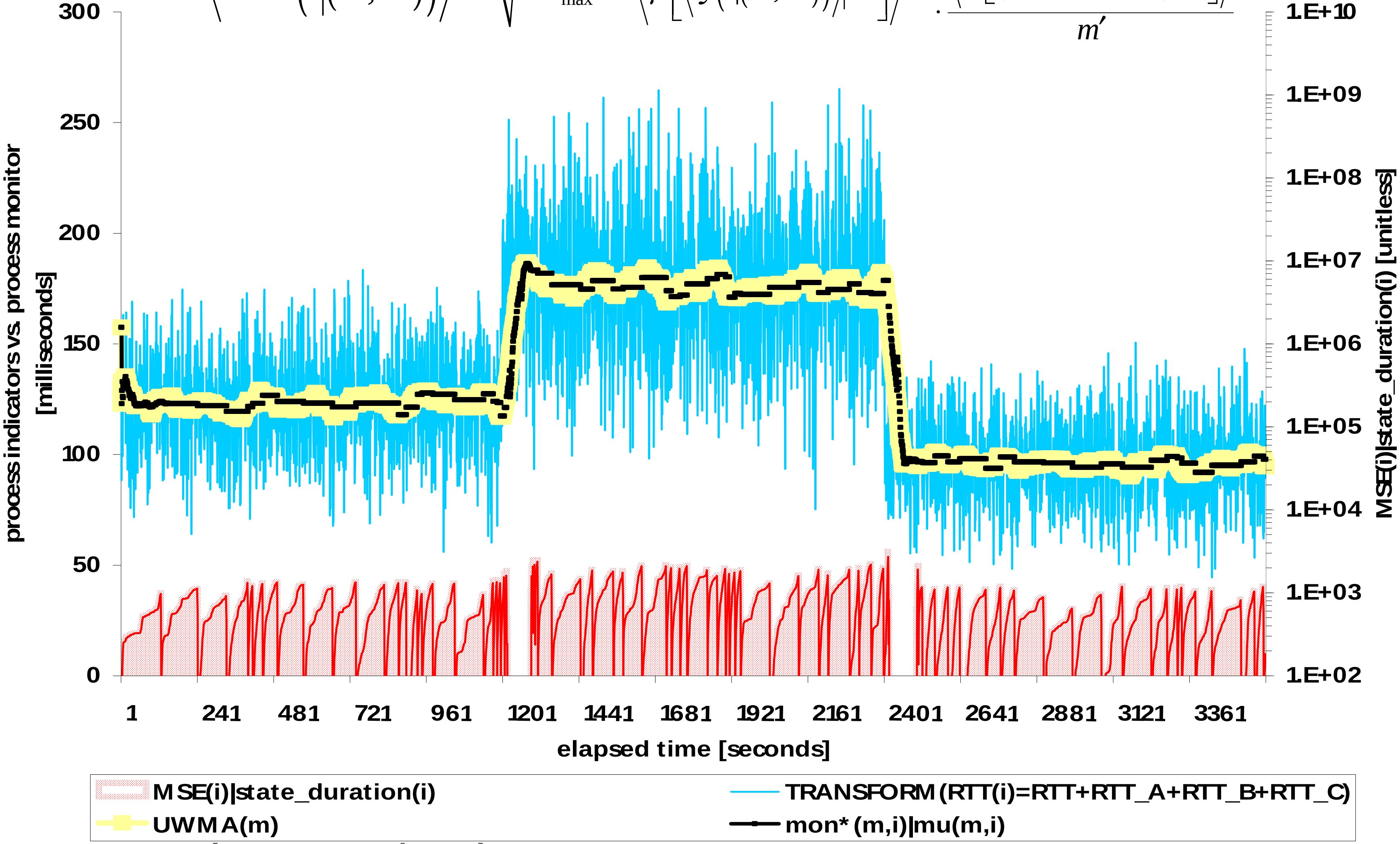
$$\left\langle MSE\left(i \middle| (m,m')\right)\right\rangle \leq \sqrt{m' \cdot t_{\max}^2 + 2 \cdot \left\langle \hat{\mu}\left[\left\langle \zeta\left(i \middle| (m,m')\right)\right\rangle \middle| m'\right]\right\rangle^2} \cdot \frac{\left\langle \hat{\mu}\left[\left\langle \hat{\sigma}_D(i) \middle| (m,m')\right\rangle \middle| m'\right]\right\rangle}{m'}$$

- is bounded by **(5.8)**, where
  - $\langle \mu[\langle \sigma_D(i) \mid (m,m') \rangle] \rangle$ is average of pooled stddev along $\langle v \rangle$, and
  - $t_{max} \equiv t(m+m'-2, \alpha/2)$.
  - $\langle \zeta(i) \mid (m,m') \rangle$ represents an error correlation **(5.9)**

$$\left\langle \zeta\left(i \middle| (m,m')\right)\right\rangle = \frac{\left\langle \hat{\varepsilon}_{fast}(i)\right\rangle \cdot \left\langle \hat{\varepsilon}_{slow}(i)\right\rangle}{\left\langle \hat{\sigma}_D(i) \middle| (m,m')\right\rangle^2}$$

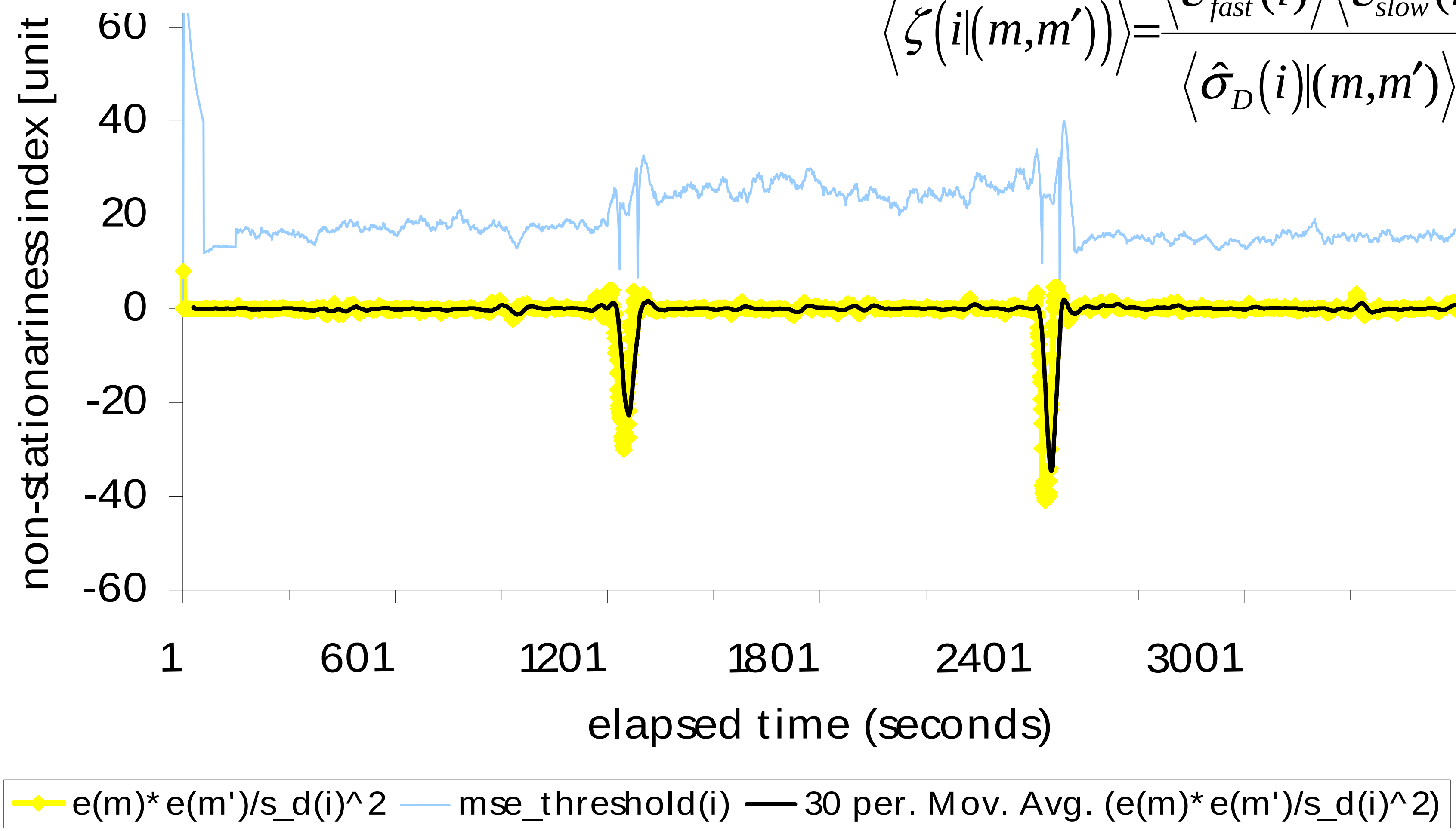
HPS MSE Equivalency Theorem:

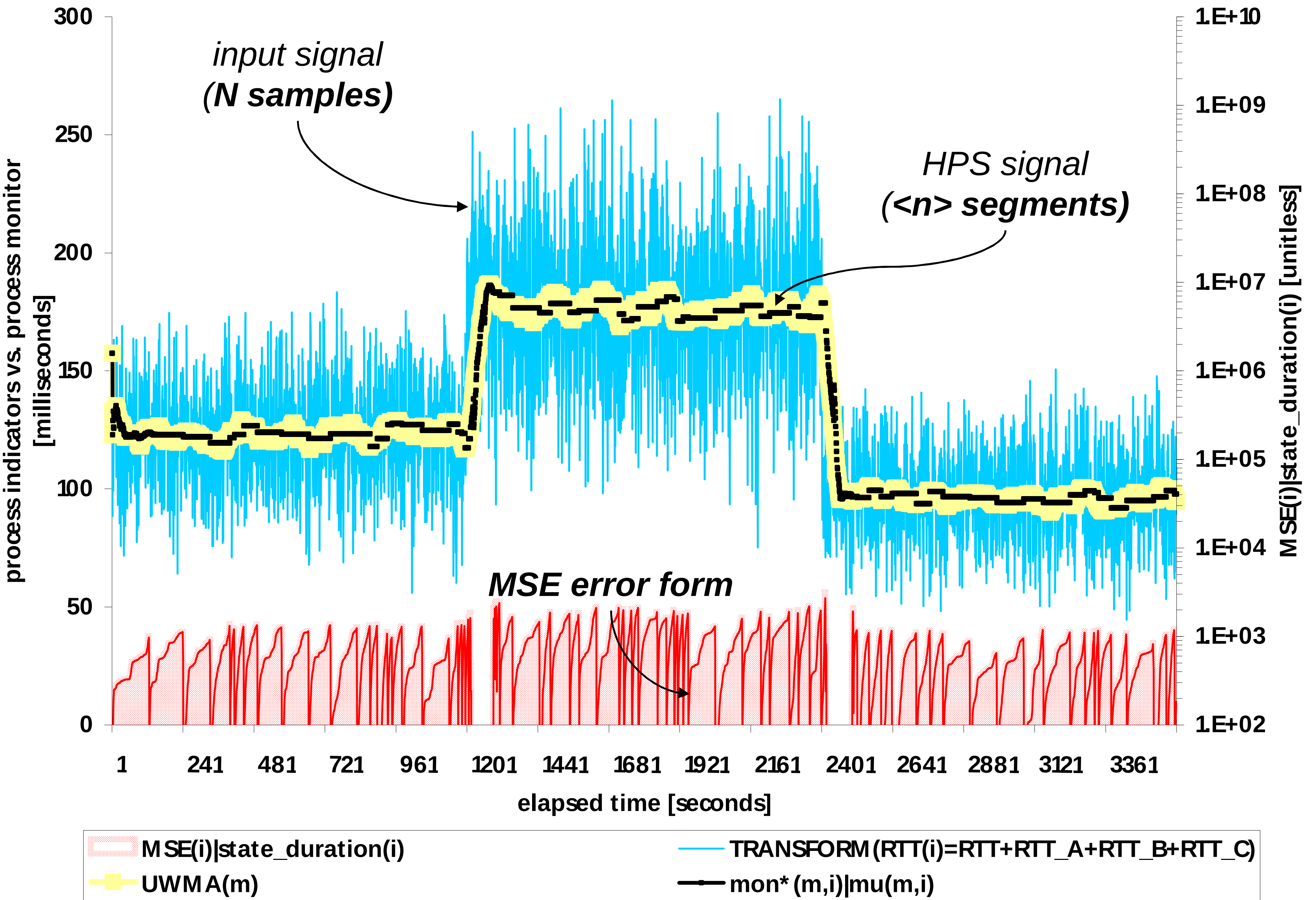
$$\left\langle MSE\left(i\middle|(m,m')\right)\right\rangle \le \sqrt{m'\cdot t_{\max}^{2}+2\cdot\left\langle \hat{\mu}\left[\left\langle \zeta\left(i|(m,m')\right)\right\rangle|m'\right]\right\rangle^{2}}\cdot\frac{\left\langle \hat{\mu}\left[\left\langle \hat{\sigma}_{D}(i)|(m,m')\right\rangle|m'\right]\right\rangle}{m'}$$



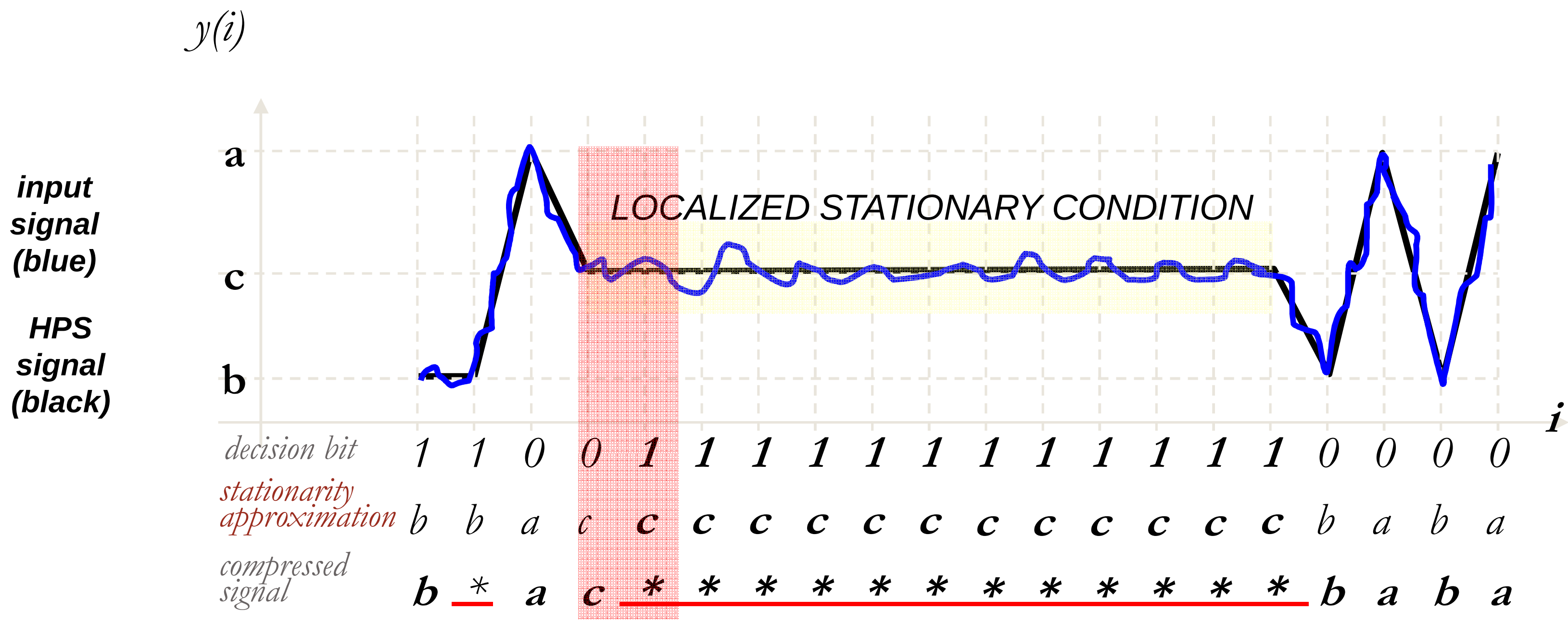
Behavior of the error correlation

$$\left\langle \zeta\left(i \| (m,m')\right)\right\rangle = \frac{\left\langle \hat{\varepsilon}_{fast}(i)\right\rangle \cdot \left\langle \hat{\varepsilon}_{slow}(i)\right\rangle}{\left\langle \hat{\sigma}_D(i) \| (m,m')\right\rangle^2}$$




input signal
(N samples)
HPS signal
(<n> segments)
MSE error form
process indicators vs. process monitor
[milliseconds]
MSE(i)|state_duration(i) [unitless]
300
250
200
150
100
50
0
1E+10
1E+09
1E+08
1E+07
1E+06
1E+05
1E+04
1E+03
1E+02
1
241
481
721
961
1201
1441
1681
1921
2161
2401
2641
2881
3121
3361
elapsed time [seconds]
MSE(i)|state_duration(i)
TRANSFORM(RTT(i)=RTT+RTT_A+RTT_B+RTT_C)
UWMA(m)
mon*(m,i)|mu(m,i)

HPS Stationary-Based Encoding: Intuition



- Generation of stationary decision bit continually detects and encodes approximate duration, value, and location of stationary conditions regardless of timescale of such

HPS Transform: Operational Region

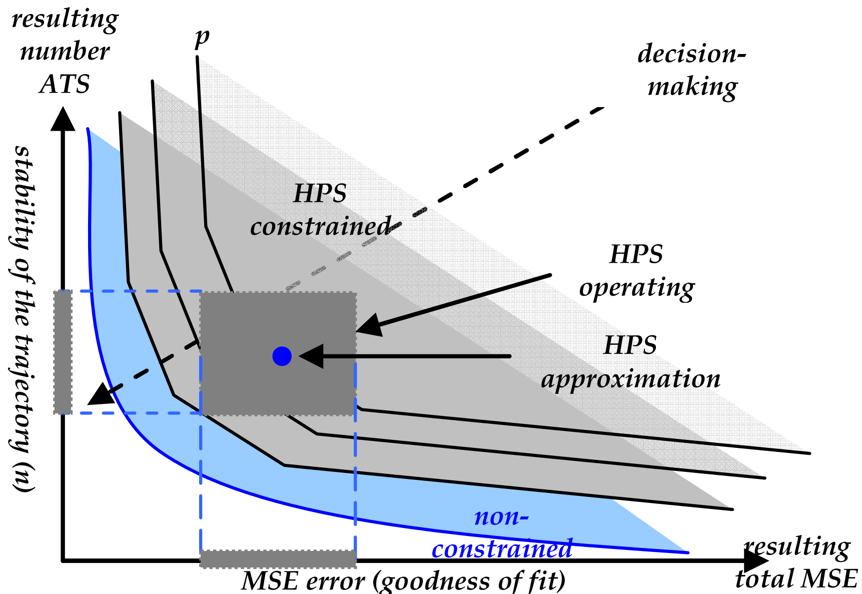


Fig. 4: Intuition into the operation region of the HPS TRANSFORM.

- Application of MSE EQUIVALENCY THEOREM transforms
  - a derivative form (two maximal likelihood estimators)
  - of an input signal of N samples
  - into a highly compressed representation of consisting of
    - **‹n›** coarse tracking-segments and **‹n›**+1 transitions
- Provides TRADEOFF control over
  - **‹n›** number of segments used to track stationary conditions
  - ***‹MSE›*** (GOODNESS-OF-FIT) of the representation

Part II:
A data reduction technique for SVM

- A data reduction technique to reduce a SVM training dataset
  - into a subsample of the training samples and
  - a careful selection of boundary conditioning points (samples that represent outliers for features)
  - while increasing prediction accuracy but lesser but carefully selected training samples

SVM is a margin classifier

- Datasets can be viewed as consisting of tuples which
  - Combine into one or more clusters within the space spanned by a feature or feature set
  - Do not combine into clusters
- Support vectors can come from either set but
  - A data reduction technique could be applied to the former
- Idea:
  - For each feature, find outliers with respect to population of feature values
  - Decompose dataset into two different sets with respect to the presence of outliers
  - Recombine the sub-datasets into a new training dataset based on some criteria

Approach for SVM Data Reduction

- Given a training set FULL_SET
- Identify per-feature, within feature values, outliers;
  - for example, for numerical features select outliers at K sigma levels from feature mean
- Identify training samples with D (e.g., 1) or more feature outliers
- Separate training dataset into two disjoint subsets:
  - OUTLIER_SET: those samples having at least D feature outliers
  - NORMAL_SET: those samples having less that D feature outliers
- Subsample NORMAL_SET by some fraction r
- Generate an ADJUSTED_SET by mixing subsampled NORMAL_SET with OUTLIER_SET
- Train SVM with ADJUSTED_SET

Performance Numbers (In Progress)

- Preliminary data obtained for yahoo answers dataset so far indicate increase in prediction accuracy (at K=3 sigma levels for extracting OUTLIER_SET) from
  - 72 % accuracy to
  - 75 % accuracy @ 50 % subsampling over NORMAL_SET
  - 80 % accuracy @ 65% subsampling over NORMAL_SET
  - 88 % accuracy @ 75% subsampling over NORMAL_SET
- For datasets with low outlier density,
  - no increase in performance is observed
  - OUTLIER_SET can be nil and consequently, subsampling of NORMAL_SET can actually decrease performance
- Standard SVM training and prediction datasets found from LIBSVM (w1a, svmguide, etc.) will follow in subsequent document