Future of Surrogates and Other Acceleration Techniques

Michael Pürrer, AEI Potsdam-Golm

PAX workshop, Amsterdam, Aug 14, 2017

MAX-PLANCK-GESELLSCHAFT

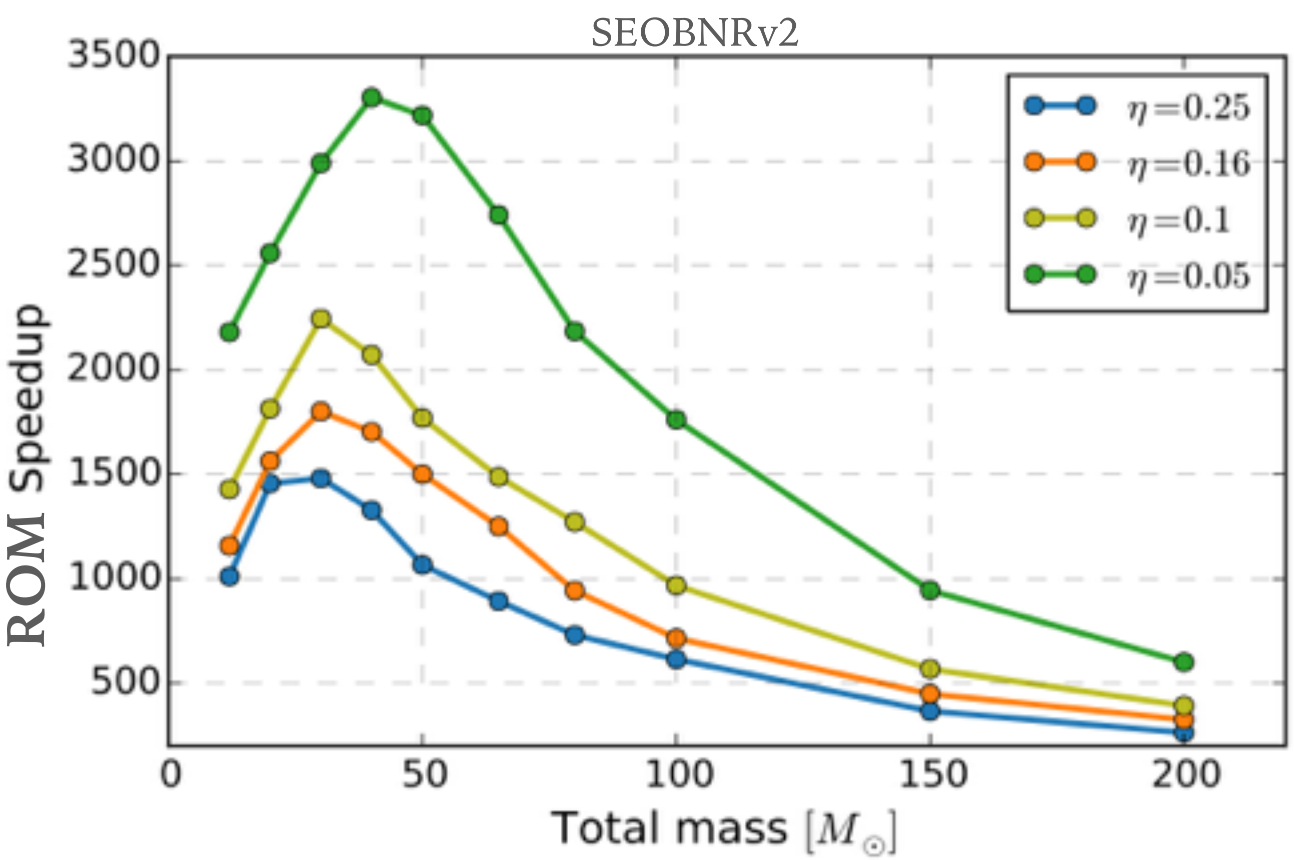
WHY SURROGATES?

- Main drivers: Parameter estimation, stoch. template banks
- Reduce **waveform / likelihood evaluation costs!**
  - **Surrogates** or **Reduced order models (ROMs)** of waveforms
  - **Reduced order quadratures (ROQs)** of likelihood
- Basic idea: replace an **existing model** with a **surrogate model**
  - surrogate is *faster* to evaluate
  - surrogate is *"close"* to original model
  - surrogate *reduces the complexity* of the original model

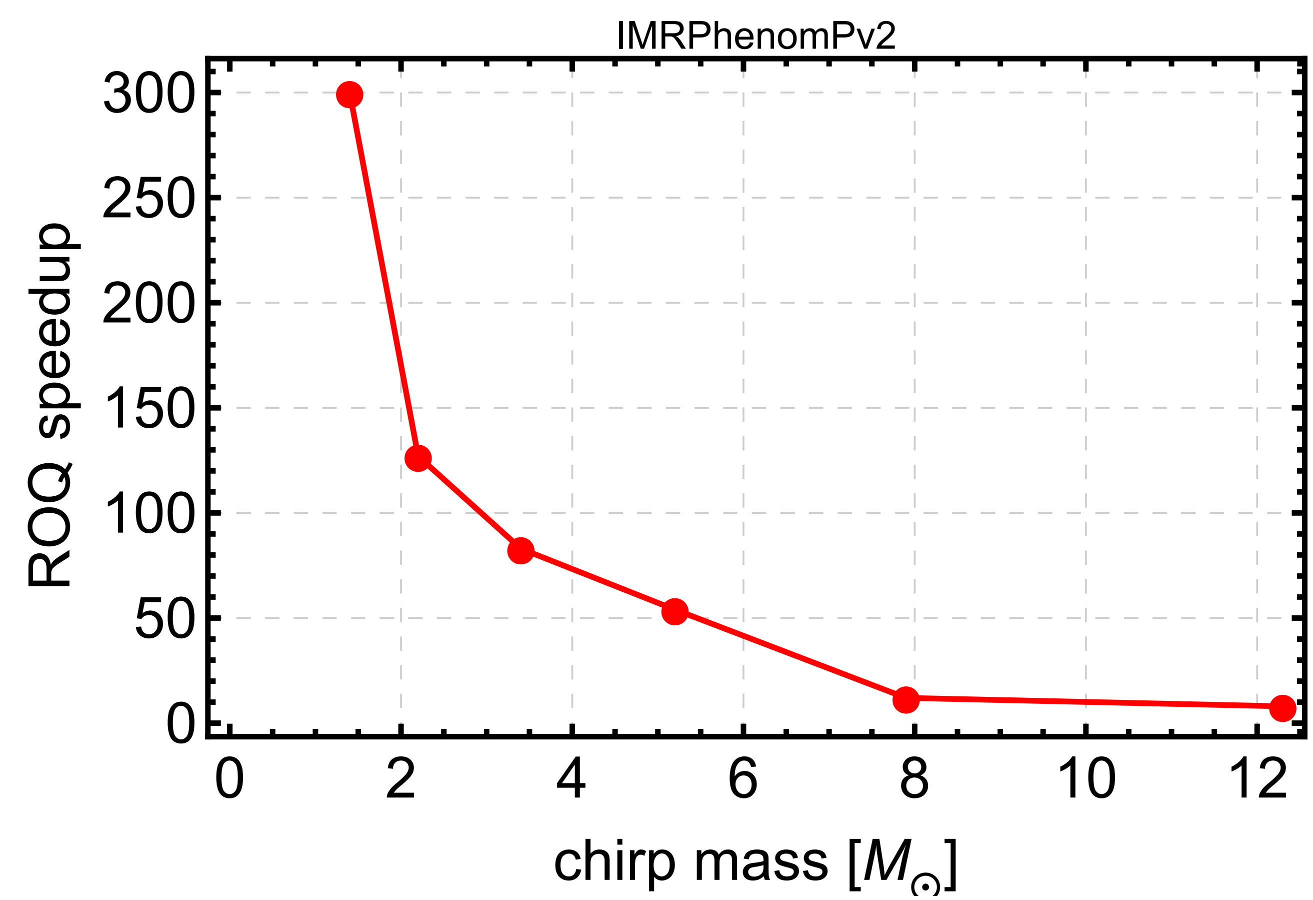
➡ A2: measurement and interpretation

REDUCING WAVEFORM COSTS

- Main drivers: Parameter estimation, stoch. template banks
- Reduce **waveform / likelihood evaluation costs!**
  - **Surrogates** or **Reduced order models (ROMs)** of waveforms
  - **Reduced order quadratures (ROQs)** of likelihood



MP arXiv:1512.02248



Smith+ arXiv:1604.08253

STRATEGIES FOR BUILDING SURROGATES

- **Basic ingredients** shared by different approaches:
  1. Generate **fine sampling** of the model over parameter space
  2. **Orthonormal bases** e_i calculated from this data (SVD, greedy)
  3. Pose **interpolation** problem for coefficients $c_i(\vec{\lambda})$
  4. **Surrogate** combines bases and interpolation data to approximate the original model
  5. **Validation** of surrogate by point-wise comparison against original model
- Steps 1, 5 *very expensive*
- Step 2 can be *expensive*

$$h(f;\vec{\lambda}) \approx \sum_{i=1}^{m} c_i(\vec{\lambda}) e_i(f)$$

$$c_i(\vec{\lambda}) = \langle h(\cdot;\vec{\lambda}), e_i(\cdot)\rangle$$

TYPES OF SURROGATES

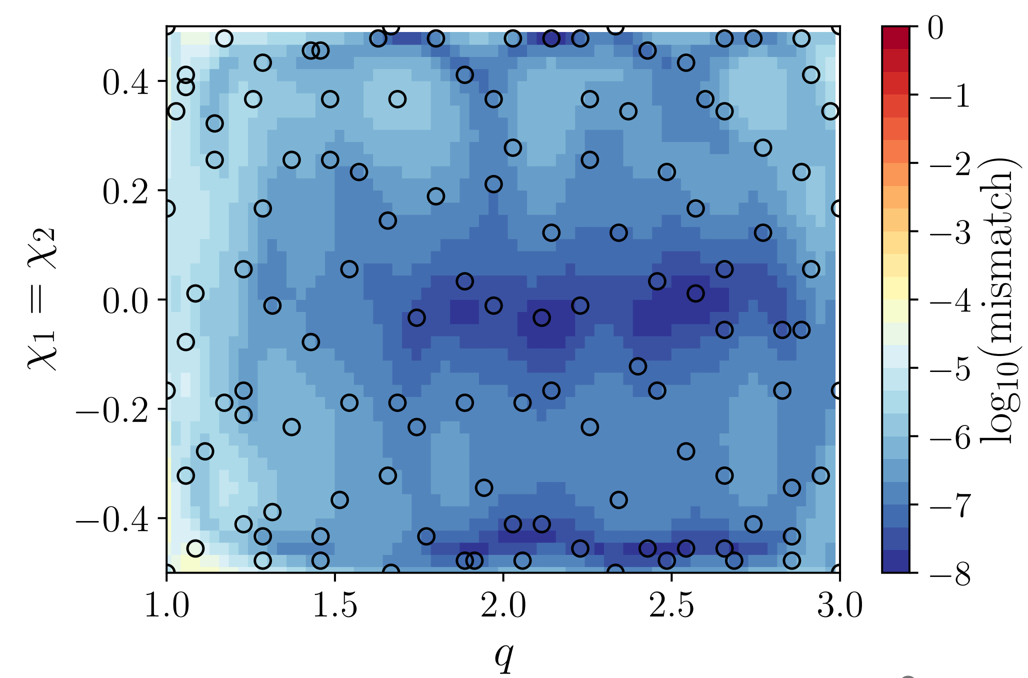
- **NR Surrogates**
  - Relatively short waveforms (~5000 M)
  - Can use dense grids (in time / frequency)
  - So far modest parameter space regions covered
  - First interpolate NR data, then build surrogate to accelerate
- **Surrogates of semi-analytic waveform models** (e.g. EOB)
  - Long waveforms to fill detector bands ($\sim 10^6$ M)
  - Requires *sparse* grids (in time / frequency)
  - Should cover parameter space where model is accurate
- **Waveform decomposition** depends on waveform morphology:
  - Aligned-spin, precession, sub-dominant modes, …

INCLUDING MODEL UNCERTAINTY

- Waveforms (NR, semi-analytic) come with various errors
- Model waveforms as *distributions* over possible functions: **Gaussian process regression (GPR)**
  - GPR can predict mean waveform + uncertainty
  - GPR can interpolate arbitrarily placed waveforms
  - Can iteratively refine a GPR-based model
- Can estimate where errors are largest
  → place new simulations there!

$$f(\vec{x}) \sim \mathcal{GP}(\vec{\mu}, \mathbf{k}(\vec{x}, \vec{x}'))$$

Mean   Covariance



Doctor+ arXiv:1706.05408

REDUCED ORDER QUADRATURE

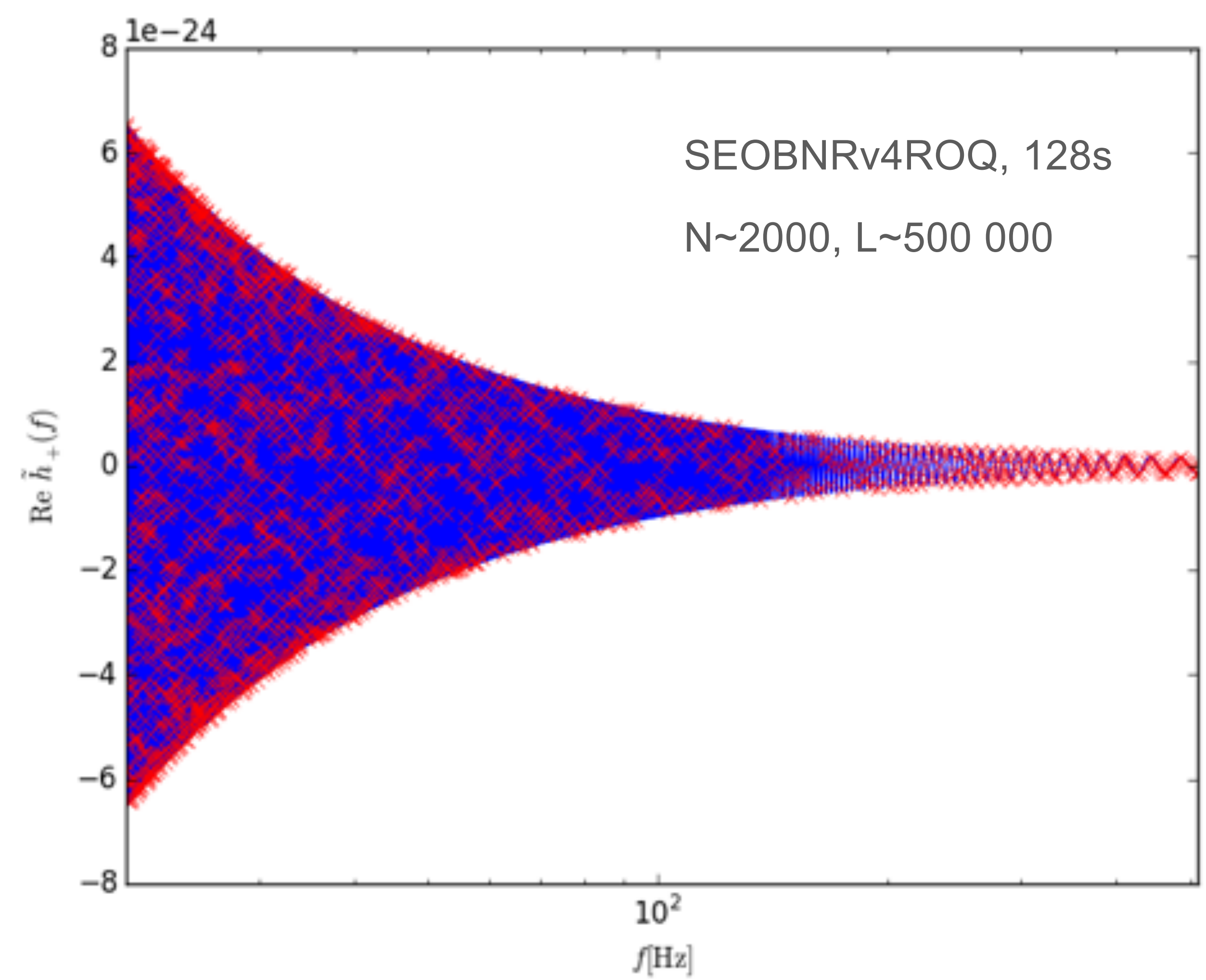
- ROQ = *weighted* numerical quadrature adapted to waveform
- *Empirical interpolant* for waveform polarizations:

$$\tilde{h}_{+,\times}(f_i;\vec{\lambda}) \approx \sum_{j=1}^{N} B_j(f_i)\tilde{h}_{+,\times}(F_j;\vec{\lambda})$$

- ROQ inner product:

$$(d, h_{+,\times}(\vec{\lambda}))_{\mathrm{ROQ}} \approx \sum_{j=1}^{N} \omega(t_c)\tilde{h}_{+,\times}(F_j;\vec{\lambda})$$

$$\omega(t_c) = 4\Re\Delta f \sum_{i=1}^{L} \frac{\tilde{d}^*(f_i)B_j(f_i)}{S_n(f_i)} e^{-2\pi i t_c f_i}$$



- Speedup if $N \ll L$

[Cañizares+ arXiv:1404.6284, Smith+ arXiv:1604.08253]

AVAILABLE ROMS / ROQS

- ROMs built from **numerical relativity** simulations
  - Blackman+ arXiv:1502.07758, arXiv:1701.00550, arXiv:1705.07089
- ROMs for **effective-one-body (EOB) models**
  - EOBNR: Field+ arXiv:1308.3565
  - SEOBNR: MP, arXiv:1402.4146, arXiv:1512.02248, Bohé+ arXiv:1611.03703
  - TEOB, Lackey+ arXiv:1610.04742
- ROMs using **Gaussian process regression**
  - Doctor+ arXiv:1706.05408, Moore+ arXiv:1412.3657, arXiv:1509.04066
- **ROQs**:
  - IMRPhenomP: Smith+ arXiv:1604.08253
  - LEA+ NSBH and SEOBNRv2/v4 ROQs in prep.
  - Related concept: multi-banding: Vinciguerra+ arXiv:1703.02062
- **Interfacing** surrogates mode-by-mode with PE: O'Shaughnessy+ arXiv:1701.01137

CHALLENGES FOR ROMS

- **Accurate interpolation** over high dimensional (~5D) parameter spaces
  - "Curse of dimensionality": points in regular grid increase exponentially with dimension
  - Interpolation of sparse data is much less accurate
  - **Precession:** Can use *spin dynamics* from *PN precession equations* to alleviate this: Blackman + arXiv:1705.07089
  - For accurate surrogates source models have to be **smooth** in parameter space
- **Fourier domain surrogates:**
  - Precession requires approximation of Wigner rotations (speed vs accuracy)
- **NR surrogates**:
  - NR simulations are of limited length: *hybridization* (before / after surrogate?)
  - How many NR waveforms are needed?
- How **accurate** do models really **have to be** for projected detector sensitivities?
  - Measurability of parameters, selection effects (which events will we see)
- ROQs: **basis size explodes** if detector sensitivity starts at low frequencies