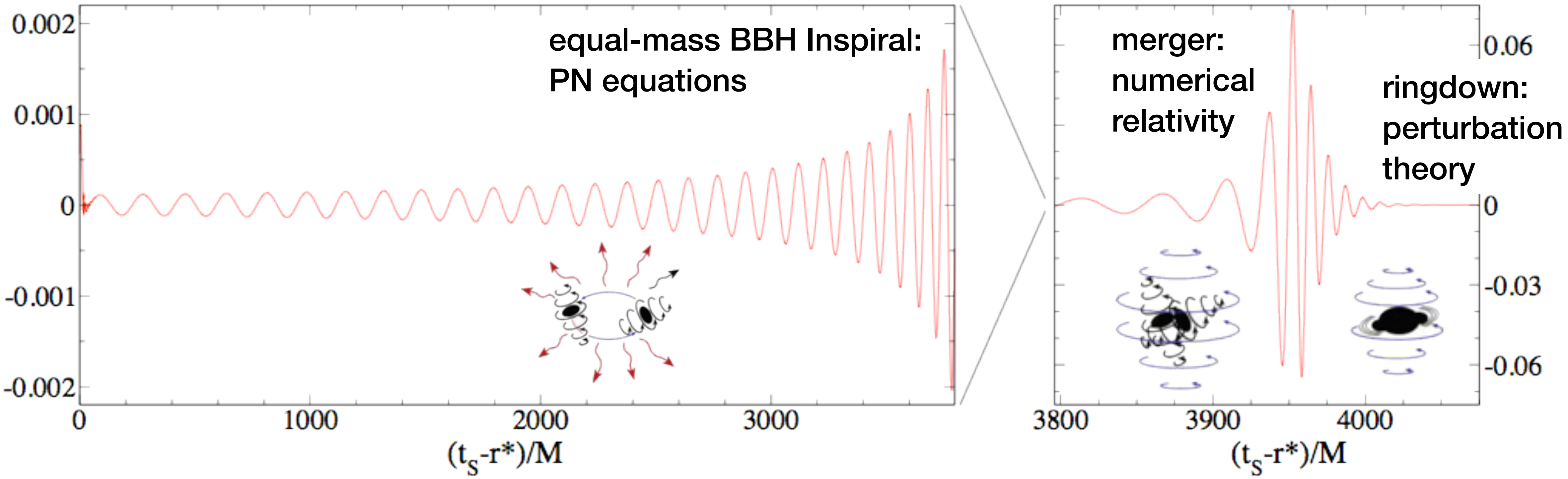
Markov Chain Monte Carlo: the ultimate multitool

Michele Vallisneri, Jet Propulsion Laboratory

GW science in a nutshell: what's in a waveform?



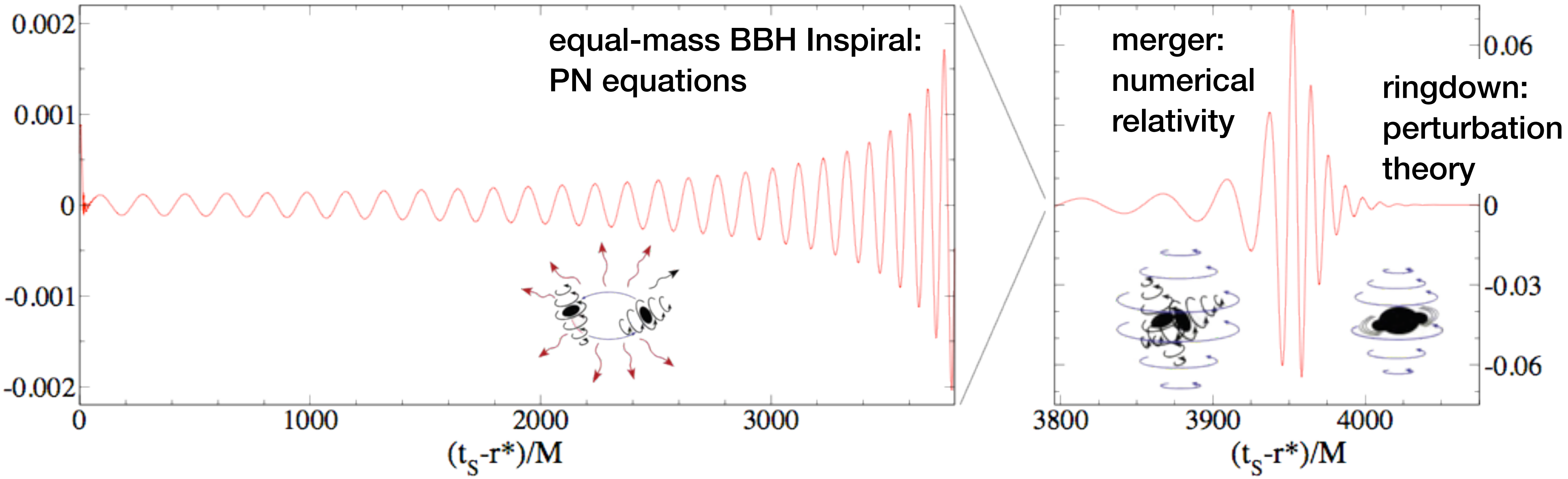
Caltech/Cornell/CITA NR

equal-mass BBH Inspiral: PN equations

merger: numerical relativity

ringdown: perturbation theory

GW science in a nutshell: what's in a waveform?



equal-mass BBH Inspiral: PN equations

merger: numerical relativity

ringdown: perturbation theory

Caltech/Cornell/CITA NR

	HF GWs: stellar masses	LF GWs: massive BHs, large separations
astrophysics	populations of compact objects; SN and GRB progenitors*	massive BH origin and evolution; Galactic WD-binary populations and interactions
nuclear physics	NS EOS, r-mode processes*	
cosmology	standard sirens*	
fundamental gravity	strong-field and radiation-sector dynamics	
black-hole structure		tests of no-hair theorem with EMRIs, ringdowns

$$p(\text{source parameters}|\text{data}) = \frac{p(\text{d.}|\text{s.p.}) \times p(\text{s.p.})}{\int p(\text{d.}|\text{s.p.}) \times p(\text{s.p.})\, d(\text{s.p.})}$$

$$\boxed{p(\text{source parameters}|\text{data})} = \frac{p(\text{d.}|\text{s.p.}) \times p(\text{s.p.})}{\int p(\text{d.}|\text{s.p.}) \times p(\text{s.p.})\, d(\text{s.p.})}$$

posterior probability:
astrophysical insight

likelihood: prob. that
noise = signal – model

$$\boxed{p(\text{source parameters}|\text{data})} = \frac{\boxed{p(\text{d.}|\text{s.p.})} \times p(\text{s.p.})}{\int p(\text{d.}|\text{s.p.}) \times p(\text{s.p.})\, d(\text{s.p.})}$$

posterior probability:
astrophysical insight

likelihood: prob. that
noise = signal – model

astrophysical
parameter prior

$$\boxed{p(\text{source parameters}|\text{data})} = \frac{\boxed{p(\text{d.}|\text{s.p.})} \times \boxed{p(\text{s.p.})}}{\int p(\text{d.}|\text{s.p.}) \times p(\text{s.p.})\, d(\text{s.p.})}$$

posterior probability:
astrophysical insight

likelihood: prob. that
noise = signal – model

astrophysical
parameter prior

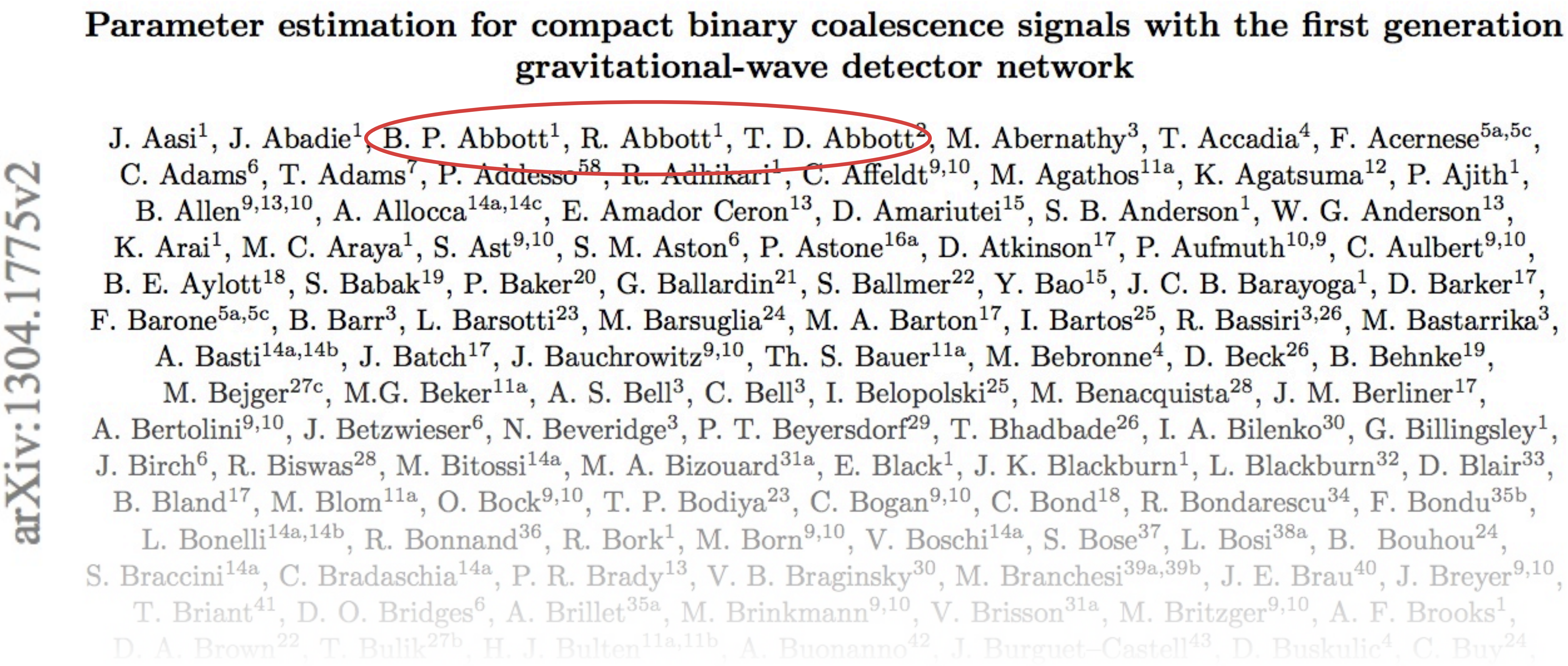
$$\boxed{p(\text{source parameters}|\text{data})} = \frac{\boxed{p(\text{d.}|\text{s.p.})} \times \boxed{p(\text{s.p.})}}{\boxed{\int p(\text{d.}|\text{s.p.}) \times p(\text{s.p.})\, d(\text{s.p.})}}$$

posterior probability:
astrophysical insight

evidence: relative prob.
that the model is right

Parameter estimation for compact binary coalescence signals with the first generation gravitational-wave detector network

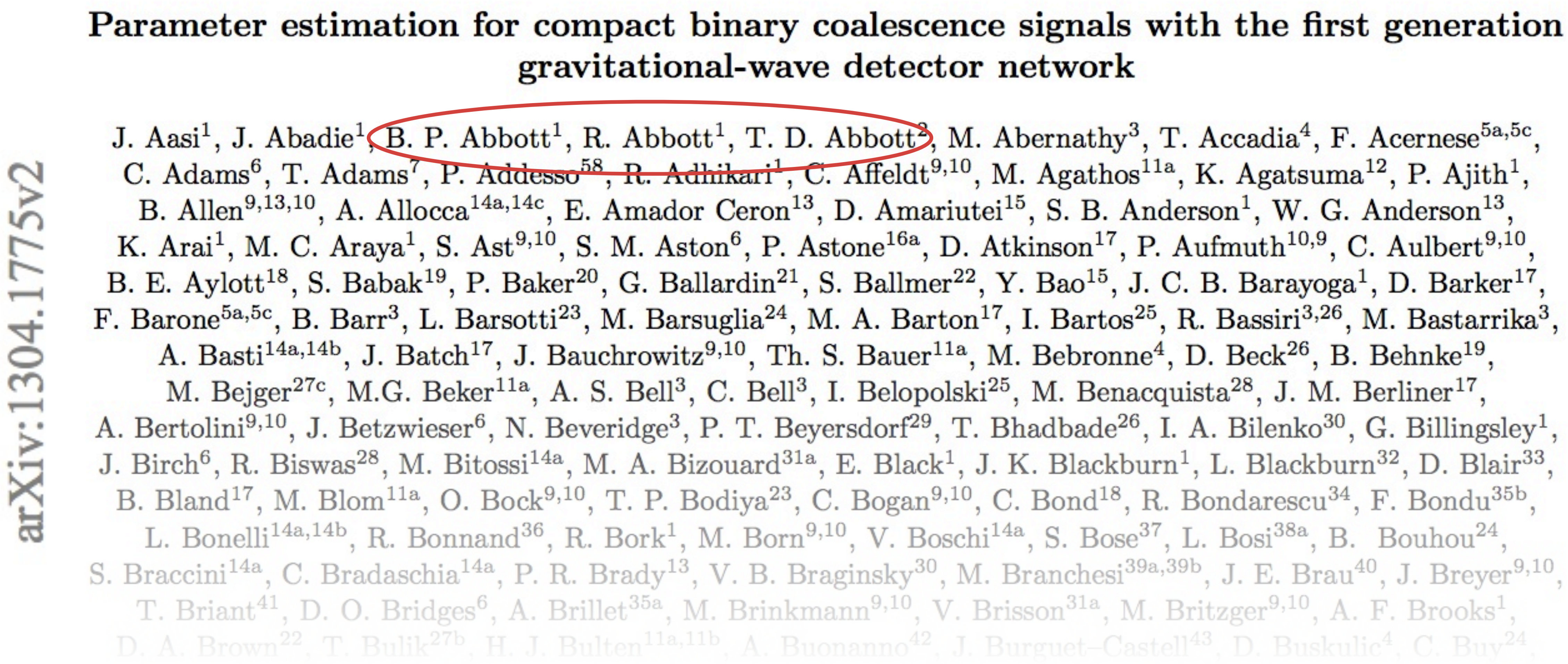
J. Aasi[1], J. Abadie[1], B. P. Abbott[1], R. Abbott[1], T. D. Abbott[2], M. Abernathy[3], T. Accadia[4], F. Acernese[5a,5c], C. Adams[6], T. Adams[7], P. Addesso[58], R. Adhikari[1], C. Affeldt[9,10], M. Agathos[11a], K. Agatsuma[12], P. Ajith[1], B. Allen[9,13,10], A. Allocca[14a,14c], E. Amador Ceron[13], D. Amariutei[15], S. B. Anderson[1], W. G. Anderson[13], K. Arai[1], M. C. Araya[1], S. Ast[9,10], S. M. Aston[6], P. Astone[16a], D. Atkinson[17], P. Aufmuth[10,9], C. Aulbert[9,10], B. E. Aylott[18], S. Babak[19], P. Baker[20], G. Ballardin[21], S. Ballmer[22], Y. Bao[15], J. C. B. Barayoga[1], D. Barker[17], F. Barone[5a,5c], B. Barr[3], L. Barsotti[23], M. Barsuglia[24], M. A. Barton[17], I. Bartos[25], R. Bassiri[3,26], M. Bastarrika[3], A. Basti[14a,14b], J. Batch[17], J. Bauchrowitz[9,10], Th. S. Bauer[11a], M. Bebronne[4], D. Beck[26], B. Behnke[19], M. Bejger[27c], M.G. Beker[11a], A. S. Bell[3], C. Bell[3], I. Belopolski[25], M. Benacquista[28], J. M. Berliner[17], A. Bertolini[9,10], J. Betzwieser[6], N. Beveridge[3], P. T. Beyersdorf[29], T. Bhadbade[26], I. A. Bilenko[30], G. Billingsley[1], J. Birch[6], R. Biswas[28], M. Bitossi[14a], M. A. Bizouard[31a], E. Black[1], J. K. Blackburn[1], L. Blackburn[32], D. Blair[33], B. Bland[17], M. Blom[11a], O. Bock[9,10], T. P. Bodiya[23], C. Bogan[9,10], C. Bond[18], R. Bondarescu[34], F. Bondu[35b], L. Bonelli[14a,14b], R. Bonnand[36], R. Bork[1], M. Born[9,10], V. Boschi[14a], S. Bose[37], L. Bosi[38a], B. Bouhou[24], S. Braccini[14a], C. Bradaschia[14a], P. R. Brady[13], V. B. Braginsky[30], M. Branchesi[39a,39b], J. E. Brau[40], J. Breyer[9,10], T. Briant[41], D. O. Bridges[6], A. Brillet[35a], M. Brinkmann[9,10], V. Brisson[31a], M. Britzger[9,10], A. F. Brooks[1], D. A. Brown[22], T. Bulik[27b], H. J. Bulten[11a,11b], A. Buonanno[42], J. Burguet–Castell[43], D. Buskulic[4], C. Buy[24],



arXiv:1304.1775v2

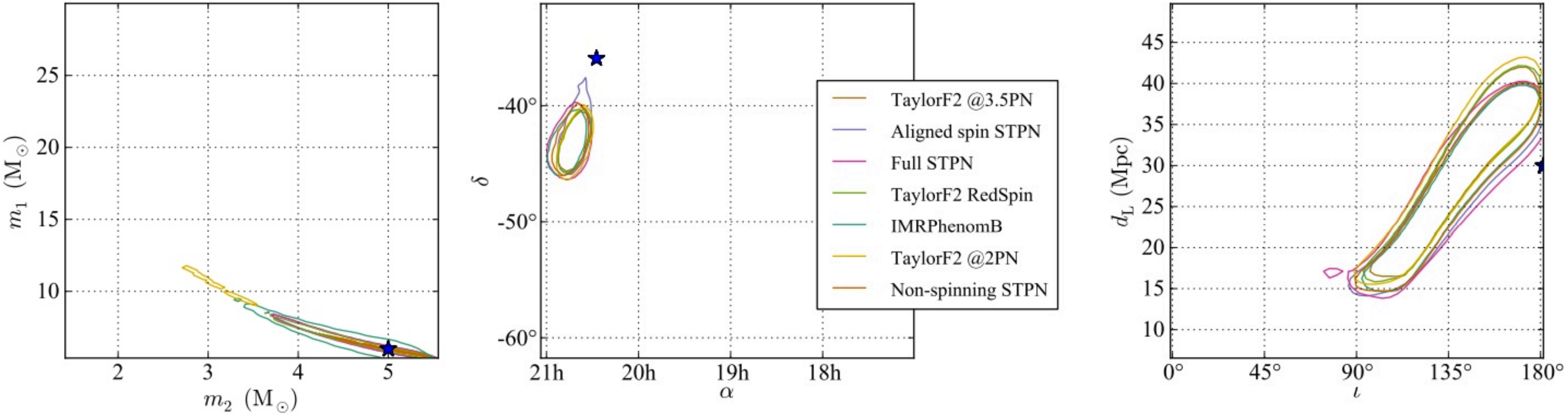
Parameter estimation for compact binary coalescence signals with the first generation gravitational-wave detector network

J. Aasi[1], J. Abadie[1], B. P. Abbott[1], R. Abbott[1], T. D. Abbott[2], M. Abernathy[3], T. Accadia[4], F. Acernese[5a,5c], C. Adams[6], T. Adams[7], P. Addesso[58], R. Adhikari[1], C. Affeldt[9,10], M. Agathos[11a], K. Agatsuma[12], P. Ajith[1], B. Allen[9,13,10], A. Allocca[14a,14c], E. Amador Ceron[13], D. Amariutei[15], S. B. Anderson[1], W. G. Anderson[13], K. Arai[1], M. C. Araya[1], S. Ast[9,10], S. M. Aston[6], P. Astone[16a], D. Atkinson[17], P. Aufmuth[10,9], C. Aulbert[9,10], B. E. Aylott[18], S. Babak[19], P. Baker[20], G. Ballardin[21], S. Ballmer[22], Y. Bao[15], J. C. B. Barayoga[1], D. Barker[17], F. Barone[5a,5c], B. Barr[3], L. Barsotti[23], M. Barsuglia[24], M. A. Barton[17], I. Bartos[25], R. Bassiri[3,26], M. Bastarrika[3], A. Basti[14a,14b], J. Batch[17], J. Bauchrowitz[9,10], Th. S. Bauer[11a], M. Bebronne[4], D. Beck[26], B. Behnke[19], M. Bejger[27c], M.G. Beker[11a], A. S. Bell[3], C. Bell[3], I. Belopolski[25], M. Benacquista[28], J. M. Berliner[17], A. Bertolini[9,10], J. Betzwieser[6], N. Beveridge[3], P. T. Beyersdorf[29], T. Bhadbade[26], I. A. Bilenko[30], G. Billingsley[1], J. Birch[6], R. Biswas[28], M. Bitossi[14a], M. A. Bizouard[31a], E. Black[1], J. K. Blackburn[1], L. Blackburn[32], D. Blair[33], B. Bland[17], M. Blom[11a], O. Bock[9,10], T. P. Bodiya[23], C. Bogan[9,10], C. Bond[18], R. Bondarescu[34], F. Bondu[35b], L. Bonelli[14a,14b], R. Bonnand[36], R. Bork[1], M. Born[9,10], V. Boschi[14a], S. Bose[37], L. Bosi[38a], B. Bouhou[24], S. Braccini[14a], C. Bradaschia[14a], P. R. Brady[13], V. B. Braginsky[30], M. Branchesi[39a,39b], J. E. Brau[40], J. Breyer[9,10], T. Briant[41], D. O. Bridges[6], A. Brillet[35a], M. Brinkmann[9,10], V. Brisson[31a], M. Britzger[9,10], A. F. Brooks[1], D. A. Brown[22], T. Bulik[27b], H. J. Bulten[11a,11b], A. Buonanno[42], J. Burguet–Castell[43], D. Buskulic[4], C. Buy[24],

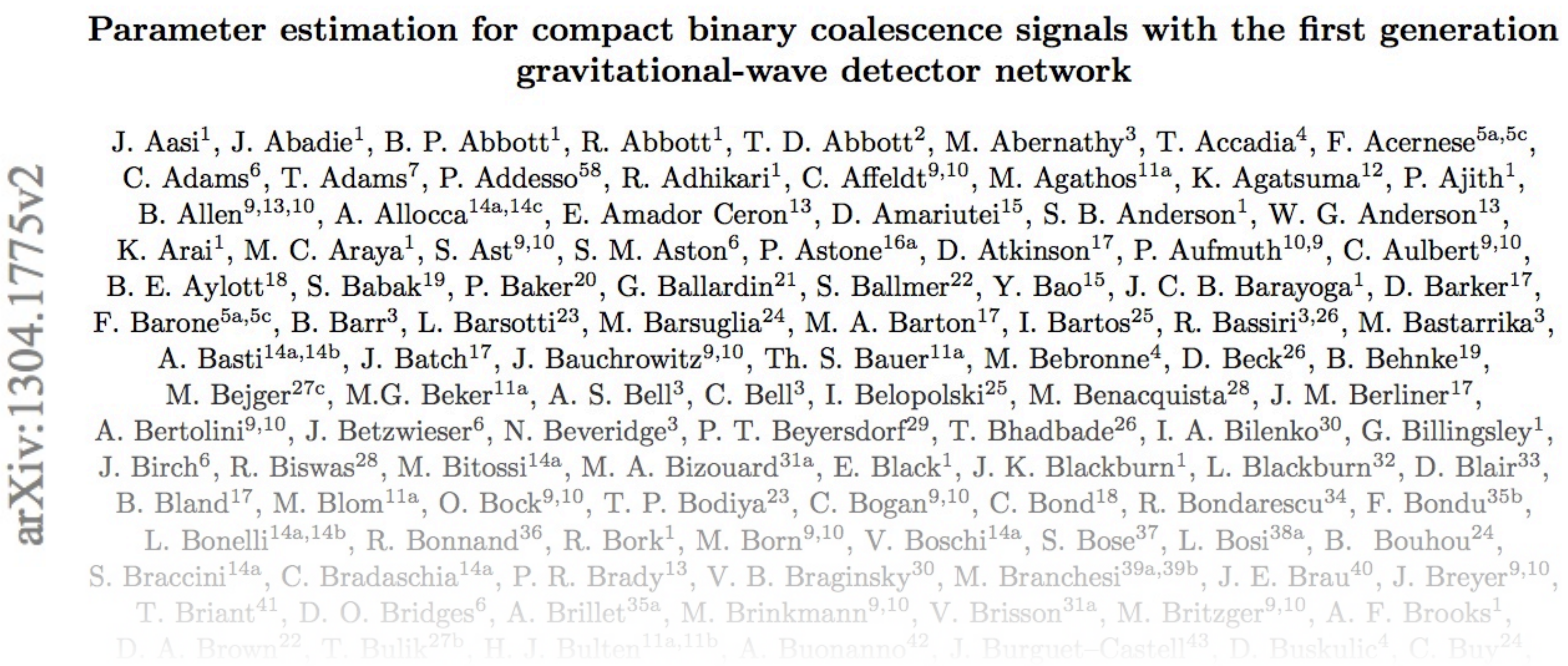


Parameter estimation for compact binary coalescence signals with the first generation gravitational-wave detector network

J. Aasi[1], J. Abadie[1], B. P. Abbott[1], R. Abbott[1], T. D. Abbott[2], M. Abernathy[3], T. Accadia[4], F. Acernese[5a,5c], C. Adams[6], T. Adams[7], P. Addesso[58], R. Adhikari[1], C. Affeldt[9,10], M. Agathos[11a], K. Agatsuma[12], P. Ajith[1], B. Allen[9,13,10], A. Allocca[14a,14c], E. Amador Ceron[13], D. Amariutei[15], S. B. Anderson[1], W. G. Anderson[13], K. Arai[1], M. C. Araya[1], S. Ast[9,10], S. M. Aston[6], P. Astone[16a], D. Atkinson[17], P. Aufmuth[10,9], C. Aulbert[9,10], B. E. Aylott[18], S. Babak[19], P. Baker[20], G. Ballardin[21], S. Ballmer[22], Y. Bao[15], J. C. B. Barayoga[1], D. Barker[17], F. Barone[5a,5c], B. Barr[3], L. Barsotti[23], M. Barsuglia[24], M. A. Barton[17], I. Bartos[25], R. Bassiri[3,26], M. Bastarrika[3], A. Basti[14a,14b], J. Batch[17], J. Bauchrowitz[9,10], Th. S. Bauer[11a], M. Bebronne[4], D. Beck[26], B. Behnke[19], M. Bejger[27c], M.G. Beker[11a], A. S. Bell[3], C. Bell[3], I. Belopolski[25], M. Benacquista[28], J. M. Berliner[17], A. Bertolini[9,10], J. Betzwieser[6], N. Beveridge[3], P. T. Beyersdorf[29], T. Bhadbade[26], I. A. Bilenko[30], G. Billingsley[1], J. Birch[6], R. Biswas[28], M. Bitossi[14a], M. A. Bizouard[31a], E. Black[1], J. K. Blackburn[1], L. Blackburn[32], D. Blair[33], B. Bland[17], M. Blom[11a], O. Bock[9,10], T. P. Bodiya[23], C. Bogan[9,10], C. Bond[18], R. Bondarescu[34], F. Bondu[35b], L. Bonelli[14a,14b], R. Bonnand[36], R. Bork[1], M. Born[9,10], V. Boschi[14a], S. Bose[37], L. Bosi[38a], B. Bouhou[24], S. Braccini[14a], C. Bradaschia[14a], P. R. Brady[13], V. B. Braginsky[30], M. Branchesi[39a,39b], J. E. Brau[40], J. Breyer[9,10], T. Briant[41], D. O. Bridges[6], A. Brillet[35a], M. Brinkmann[9,10], V. Brisson[31a], M. Britzger[9,10], A. F. Brooks[1], D. A. Brown[22], T. Bulik[27b], H. J. Bulten[11a,11b], A. Buonanno[42], J. Burguet–Castell[43], D. Buskulic[4], C. Buy[24],

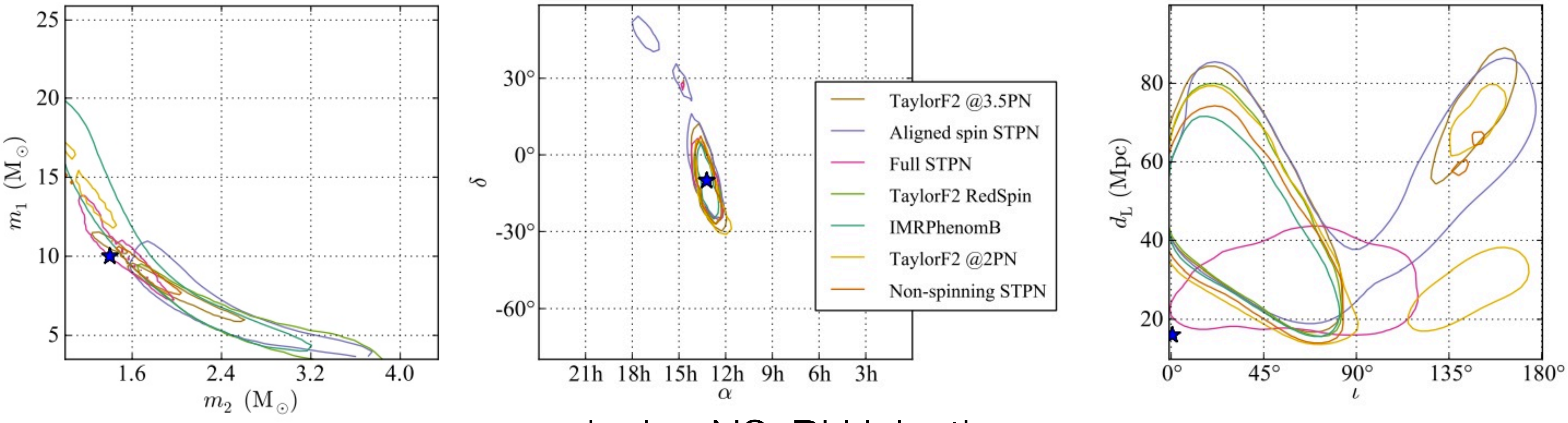


spinning BH–BH injection



Parameter estimation for compact binary coalescence signals with the first generation gravitational-wave detector network

J. Aasi[1], J. Abadie[1], B. P. Abbott[1], R. Abbott[1], T. D. Abbott[2], M. Abernathy[3], T. Accadia[4], F. Acernese[5a,5c], C. Adams[6], T. Adams[7], P. Addesso[58], R. Adhikari[1], C. Affeldt[9,10], M. Agathos[11a], K. Agatsuma[12], P. Ajith[1], B. Allen[9,13,10], A. Allocca[14a,14c], E. Amador Ceron[13], D. Amariutei[15], S. B. Anderson[1], W. G. Anderson[13], K. Arai[1], M. C. Araya[1], S. Ast[9,10], S. M. Aston[6], P. Astone[16a], D. Atkinson[17], P. Aufmuth[10,9], C. Aulbert[9,10], B. E. Aylott[18], S. Babak[19], P. Baker[20], G. Ballardin[21], S. Ballmer[22], Y. Bao[15], J. C. B. Barayoga[1], D. Barker[17], F. Barone[5a,5c], B. Barr[3], L. Barsotti[23], M. Barsuglia[24], M. A. Barton[17], I. Bartos[25], R. Bassiri[3,26], M. Bastarrika[3], A. Basti[14a,14b], J. Batch[17], J. Bauchrowitz[9,10], Th. S. Bauer[11a], M. Bebronne[4], D. Beck[26], B. Behnke[19], M. Bejger[27c], M.G. Beker[11a], A. S. Bell[3], C. Bell[3], I. Belopolski[25], M. Benacquista[28], J. M. Berliner[17], A. Bertolini[9,10], J. Betzwieser[6], N. Beveridge[3], P. T. Beyersdorf[29], T. Bhadbade[26], I. A. Bilenko[30], G. Billingsley[1], J. Birch[6], R. Biswas[28], M. Bitossi[14a], M. A. Bizouard[31a], E. Black[1], J. K. Blackburn[1], L. Blackburn[32], D. Blair[33], B. Bland[17], M. Blom[11a], O. Bock[9,10], T. P. Bodiya[23], C. Bogan[9,10], C. Bond[18], R. Bondarescu[34], F. Bondu[35b], L. Bonelli[14a,14b], R. Bonnand[36], R. Bork[1], M. Born[9,10], V. Boschi[14a], S. Bose[37], L. Bosi[38a], B. Bouhou[24], S. Braccini[14a], C. Bradaschia[14a], P. R. Brady[13], V. B. Braginsky[30], M. Branchesi[39a,39b], J. E. Brau[40], J. Breyer[9,10], T. Briant[41], D. O. Bridges[6], A. Brillet[35a], M. Brinkmann[9,10], V. Brisson[31a], M. Britzger[9,10], A. F. Brooks[1], D. A. Brown[22], T. Bulik[27b], H. J. Bulten[11a,11b], A. Buonanno[42], J. Burguet–Castell[43], D. Buskulic[4], C. Buy[24],

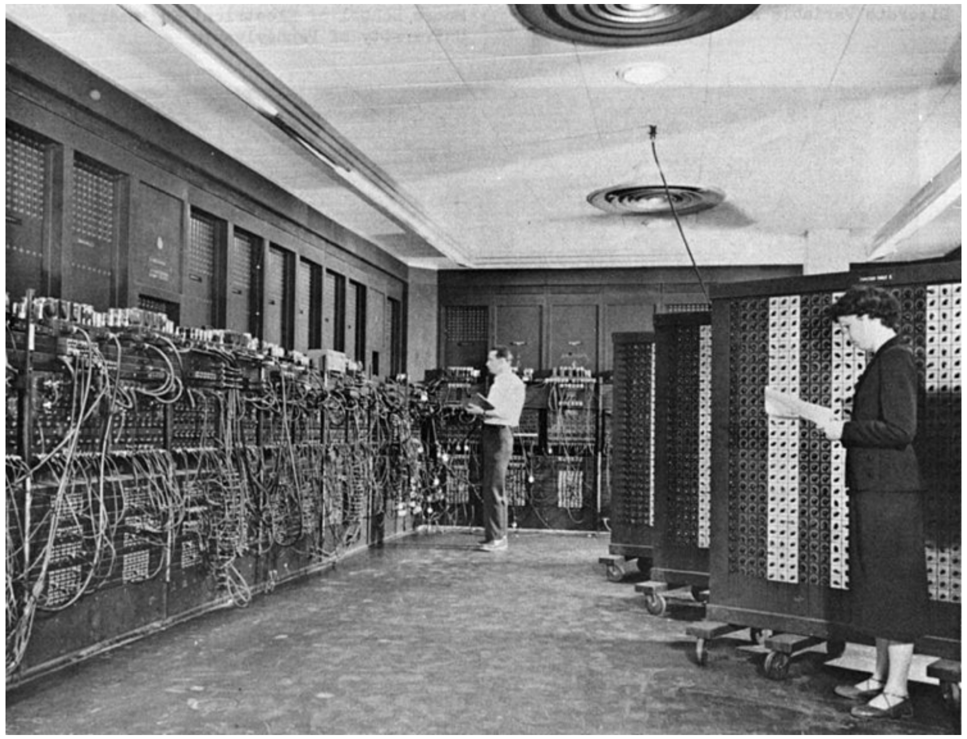


spinning NS–BH injection

Monte Carlo (Von Neumann and Ulam, 1946):
computational techniques that use random numbers

Monte Carlo (Von Neumann and Ulam, 1946):
computational techniques that use random numbers

Monte Carlo (Von Neumann and Ulam, 1946):
computational techniques that use random numbers



Monte Carlo (Von Neumann and Ulam, 1946): computational techniques that use random numbers

$$\int \phi(x)dx \rightarrow \hat{\phi} = \frac{1}{R}\sum_r \phi(x^{(r)})$$

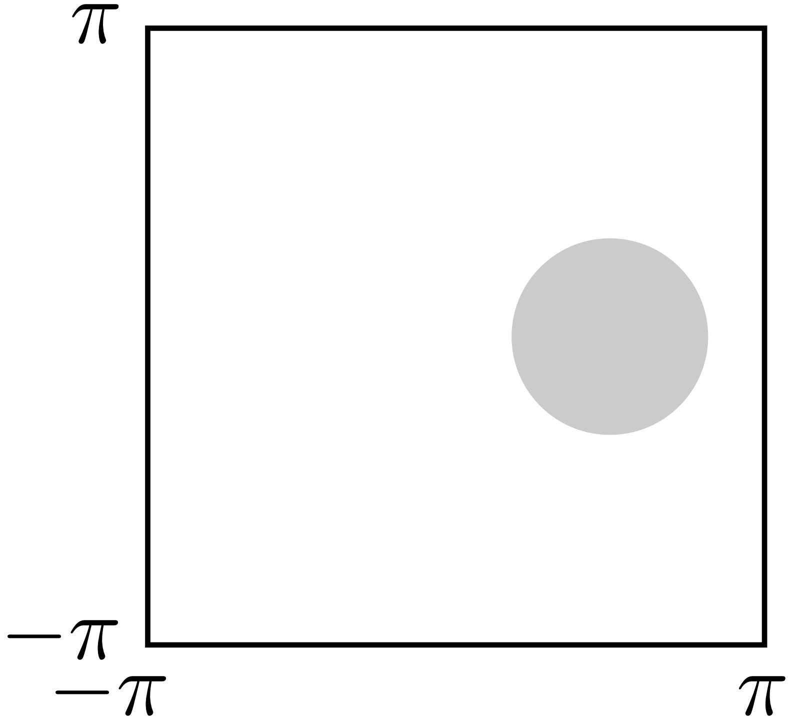
accuracy depends only on variance,
not on the number of dimensions

$$\int \phi(x)dx \rightarrow \hat{\phi} = \frac{1}{R}\sum_r \phi(x^{(r)})$$

$$\mathrm{var}\,\hat{\phi} = \frac{\mathrm{var}\,\phi}{R}$$

unfortunately uniform sampling is extremely inefficient in high-dimensional spaces

unfortunately uniform sampling is extremely inefficient in high-dimensional spaces



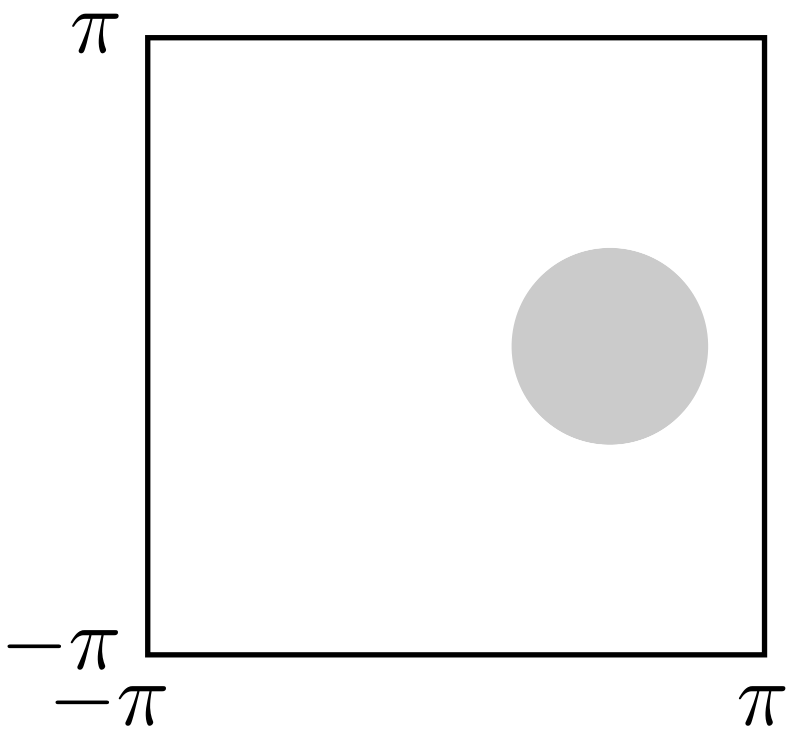
$$V_{\text{box}} = (2\pi)^d$$

$$V_{\text{ball}} = \frac{(\pi)^{d/2}}{\Gamma(n/2+1)}$$

$$\frac{V_{\text{box}}}{V_{\text{ball}}} \sim d^d$$

unfortunately uniform sampling is extremely inefficient in high-dimensional spaces

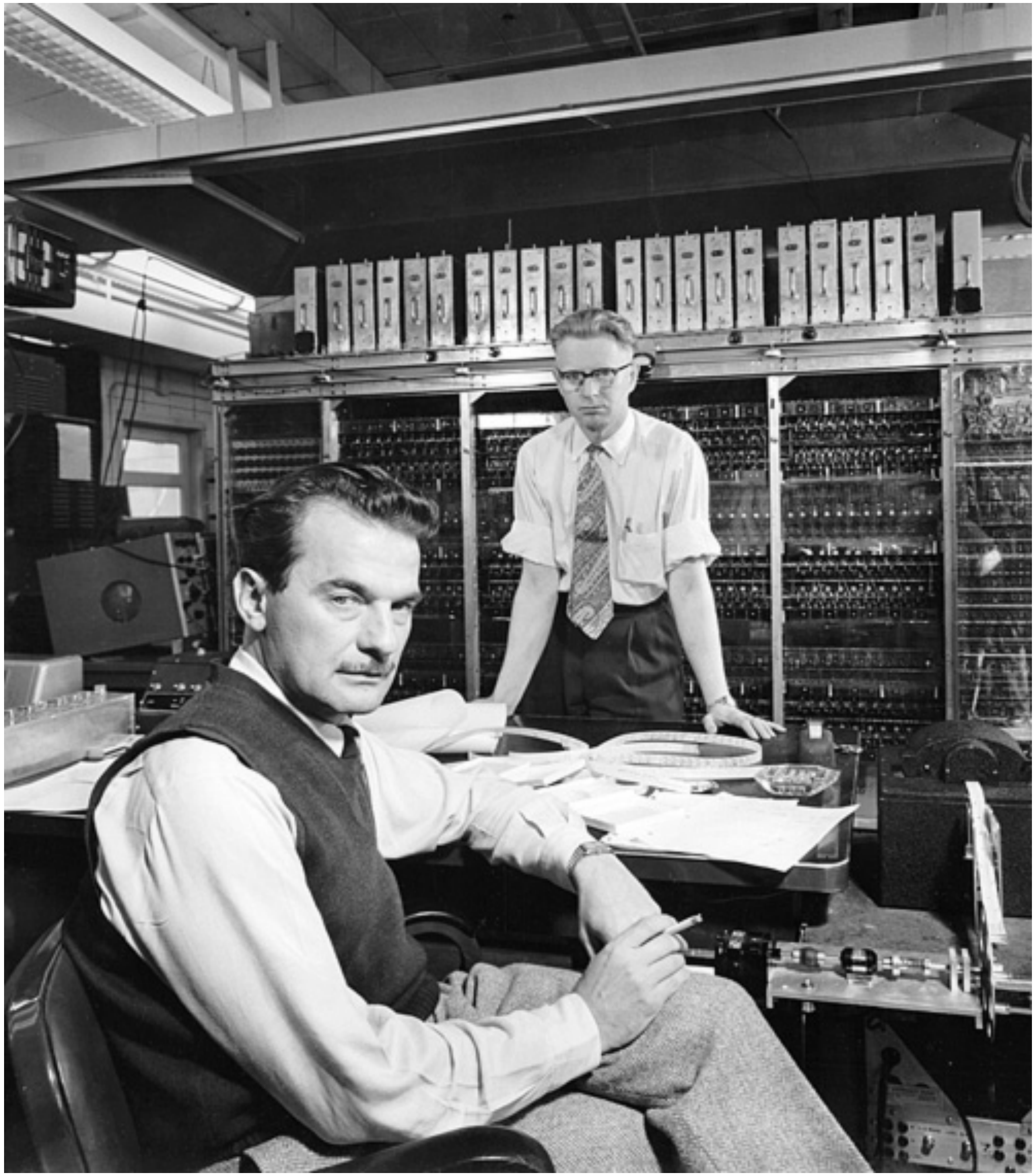
(and so are importance sampling and rejection sampling)

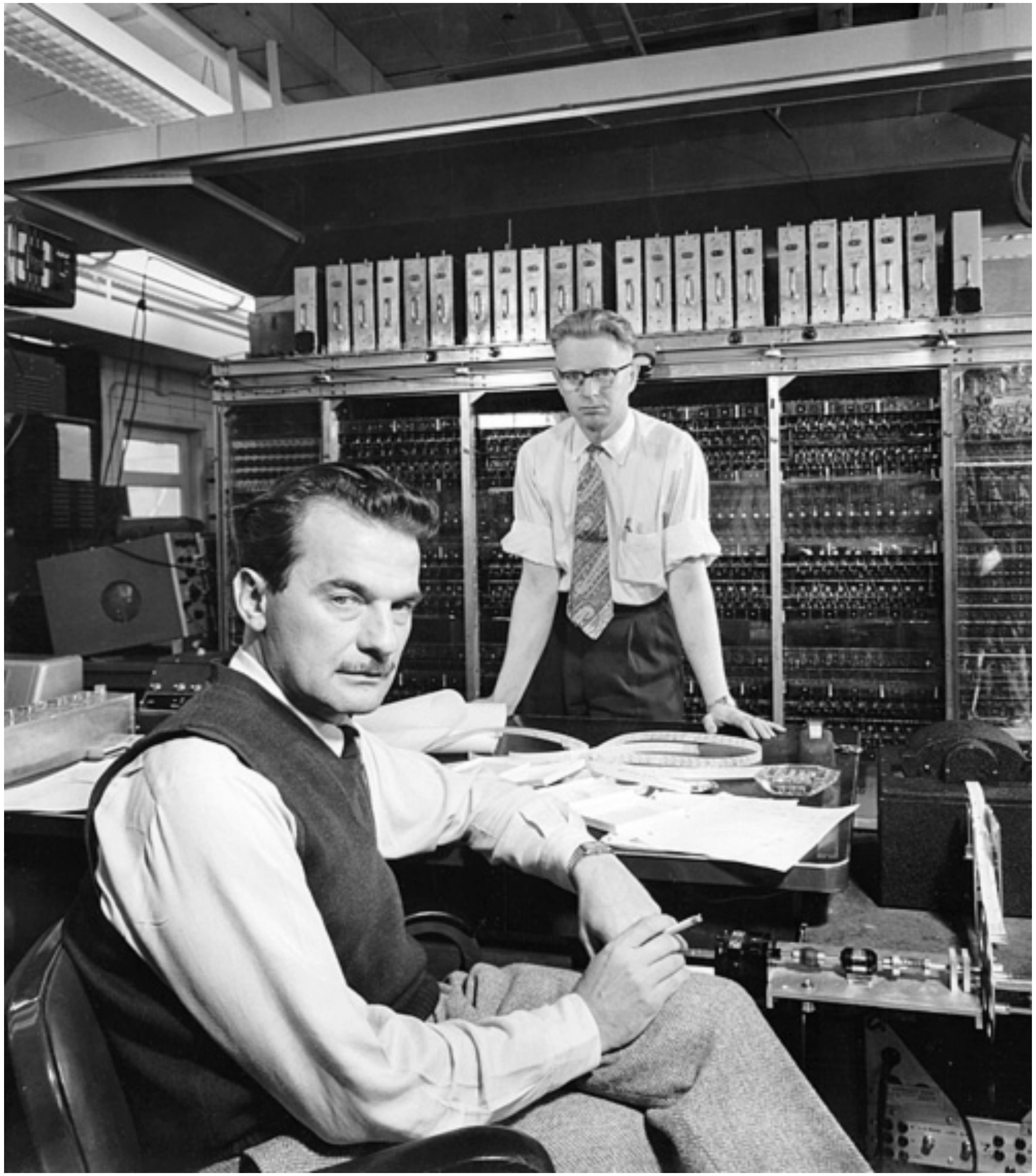


$$V_{\text{box}} = (2\pi)^d$$

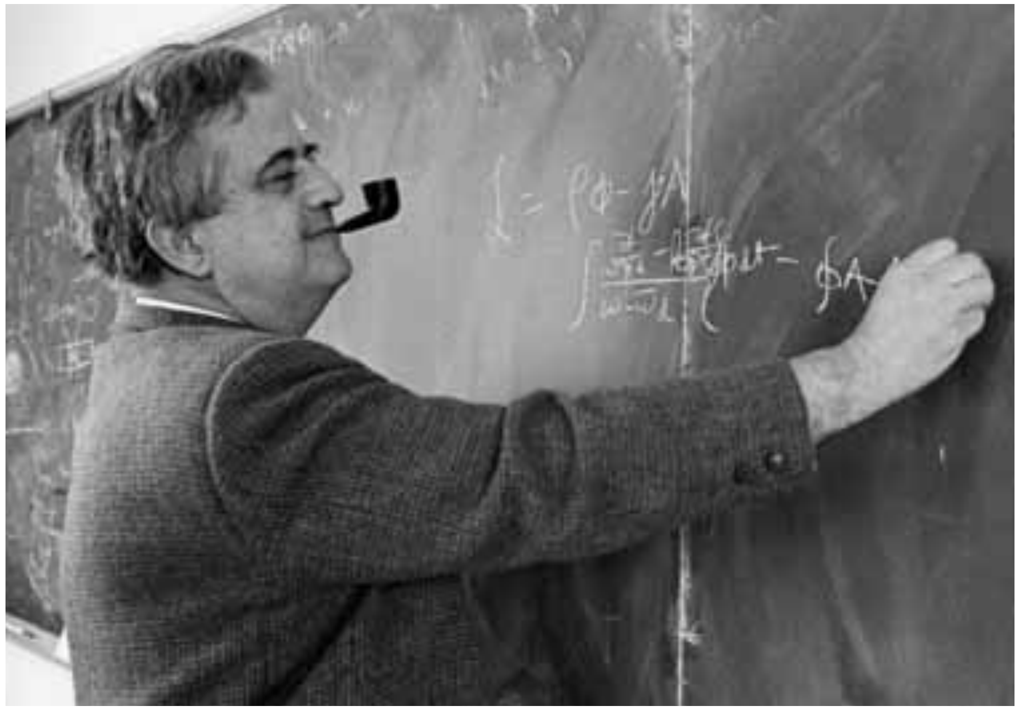
$$V_{\text{ball}} = \frac{(\pi)^{d/2}}{\Gamma(n/2+1)}$$

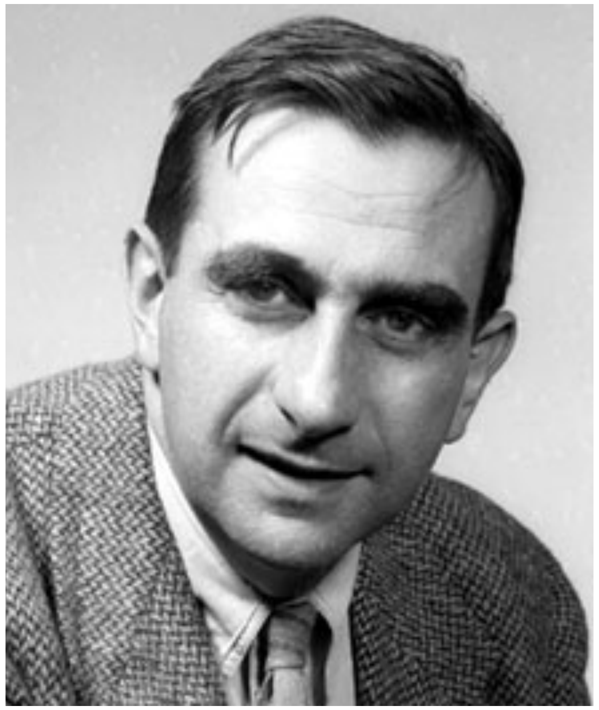
$$\frac{V_{\text{box}}}{V_{\text{ball}}} \sim d^d$$

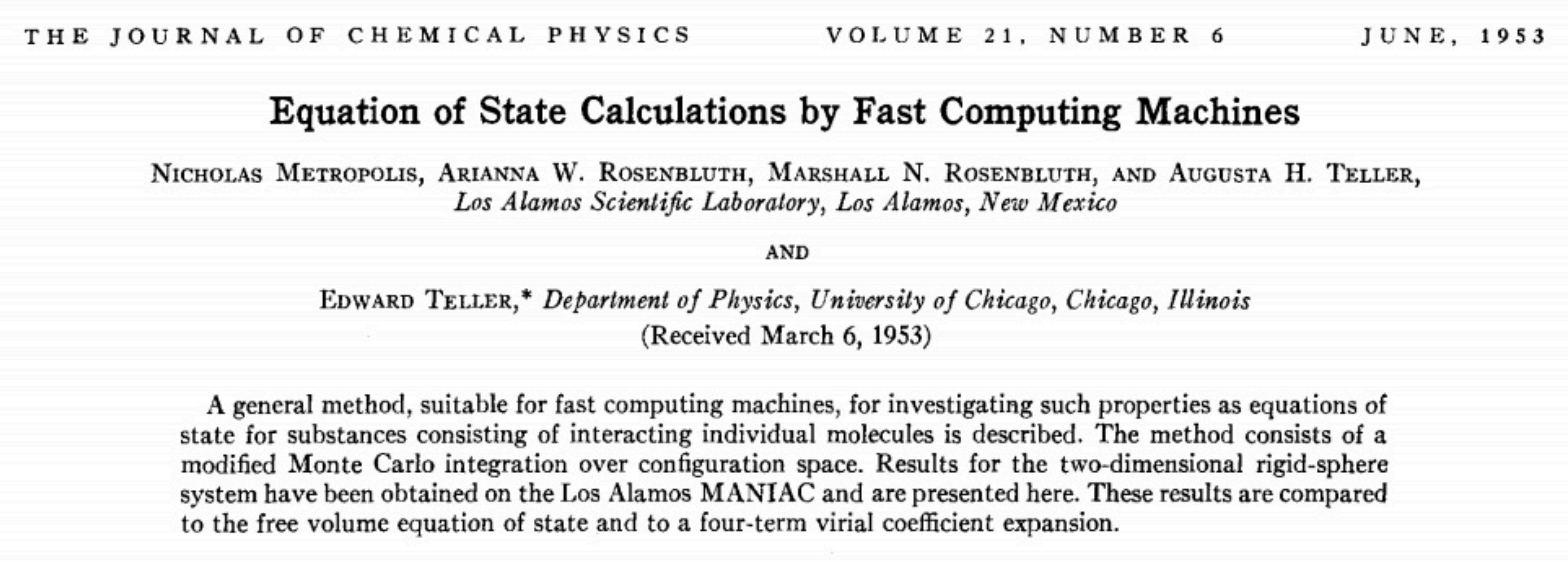




Nicholas Metropolis and his
Mathematical Analyzer Numerical Integrator And Calculator







THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

Equation of State Calculations by Fast Computing Machines

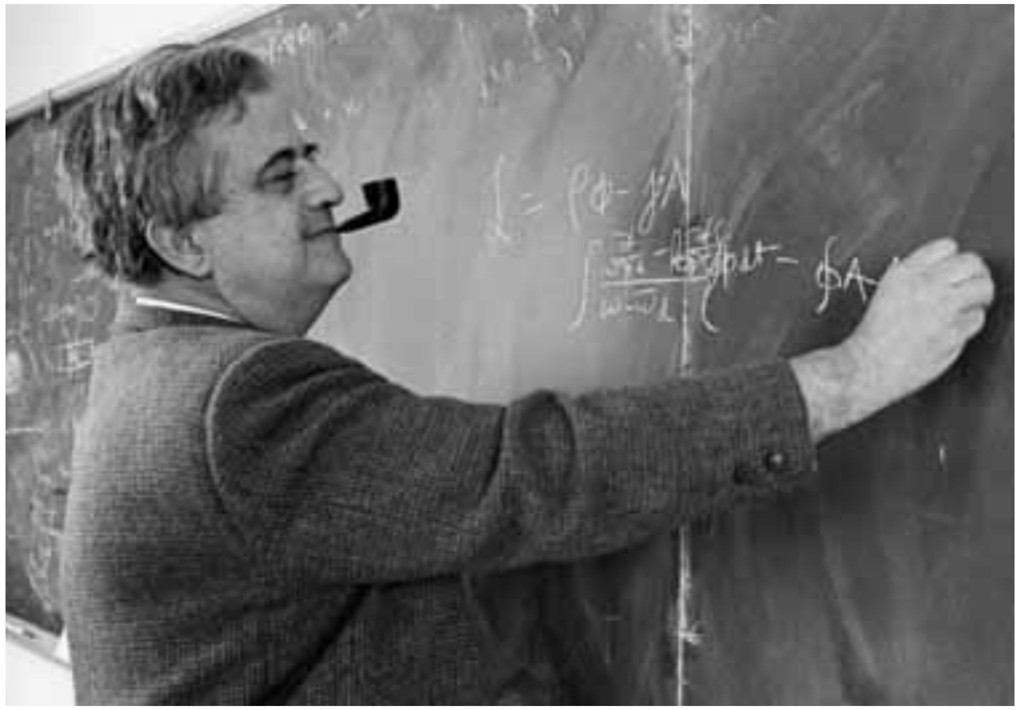
NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

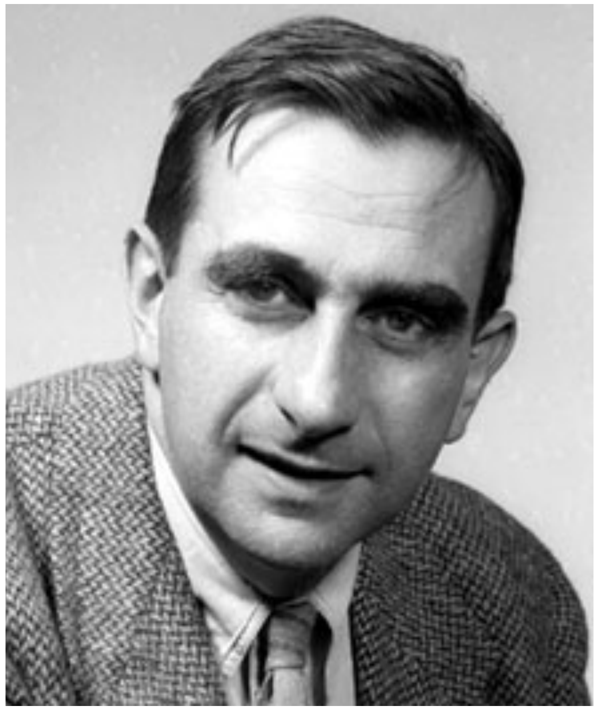
AND

EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

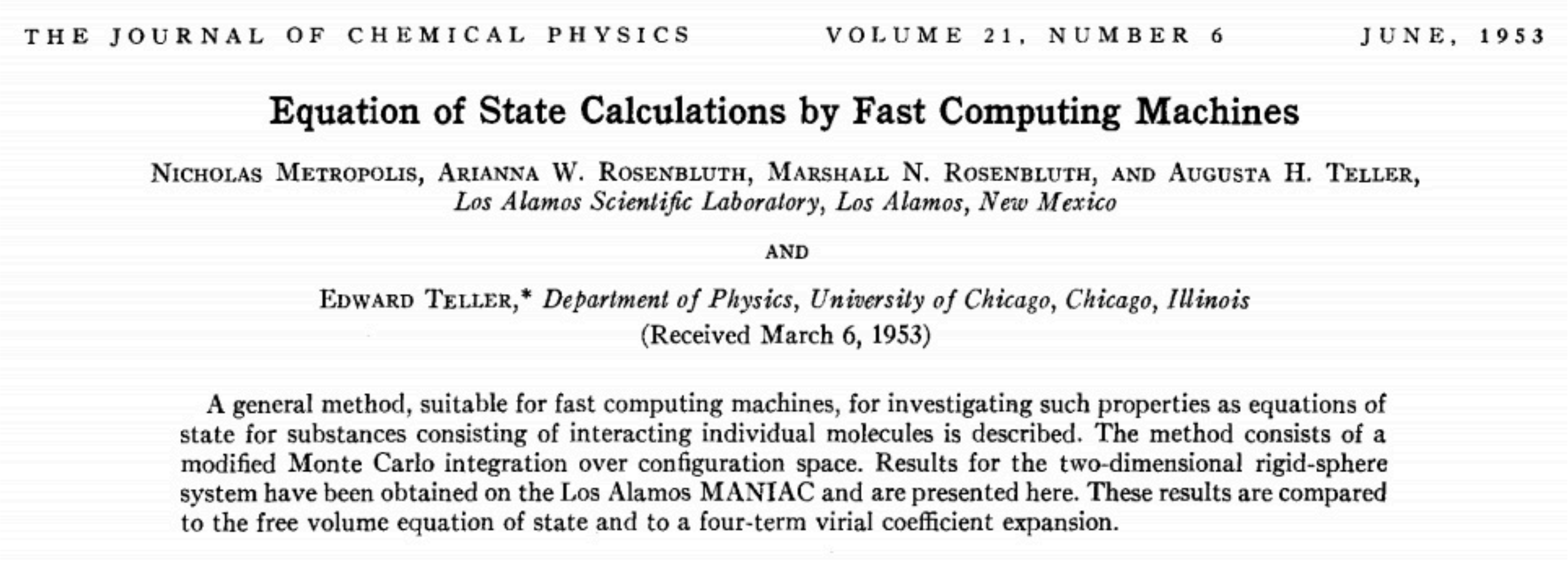
(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.





Marshall Rosenbluth and Edward Teller



THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

Equation of State Calculations by Fast Computing Machines

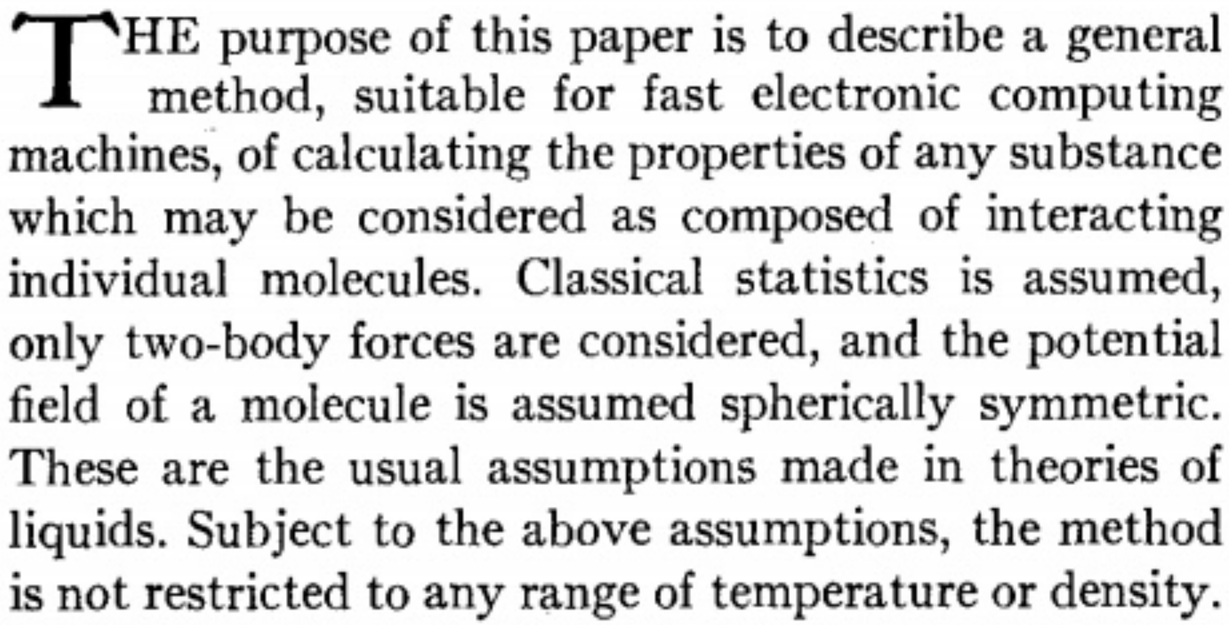
NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

AND

EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.



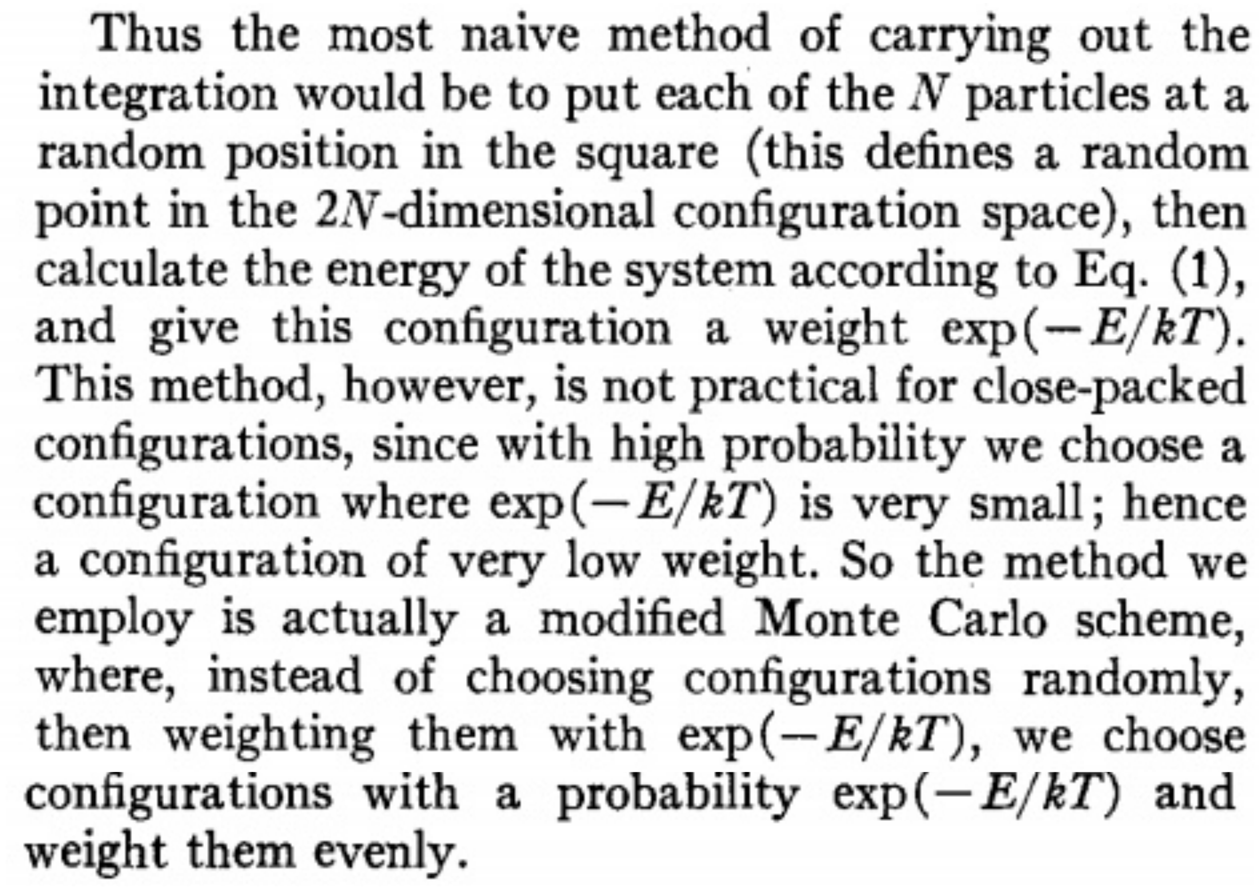
THE purpose of this paper is to describe a general method, suitable for fast electronic computing machines, of calculating the properties of any substance which may be considered as composed of interacting individual molecules. Classical statistics is assumed, only two-body forces are considered, and the potential field of a molecule is assumed spherically symmetric. These are the usual assumptions made in theories of liquids. Subject to the above assumptions, the method is not restricted to any range of temperature or density.

Teller’s crucial suggestion: ensemble averaging...

$$\int \phi(x)p(x)dx, \quad \text{with} \quad p(x) \simeq e^{-E(x)/kT}$$

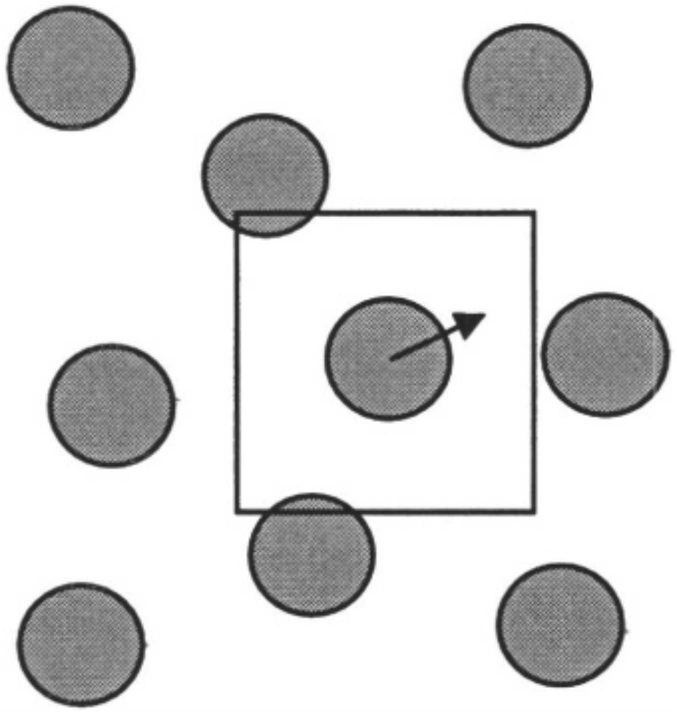
$$\Downarrow$$

$$\int \phi(x)dp(x) \simeq \frac{1}{R}\sum_R \phi(x^{(r)}) \quad \text{with} \quad \{x^{(r)}\}_P$$

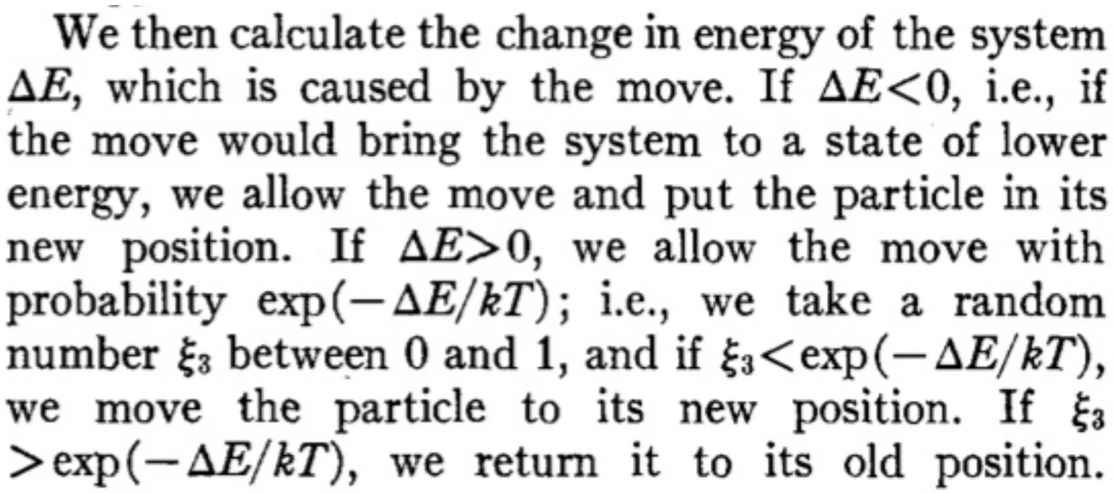


Thus the most naive method of carrying out the integration would be to put each of the N particles at a random position in the square (this defines a random point in the $2N$-dimensional configuration space), then calculate the energy of the system according to Eq. (1), and give this configuration a weight $\exp(-E/kT)$. This method, however, is not practical for close-packed configurations, since with high probability we choose a configuration where $\exp(-E/kT)$ is very small; hence a configuration of very low weight. So the method we employ is actually a modified Monte Carlo scheme, where, instead of choosing configurations randomly, then weighting them with $\exp(-E/kT)$, we choose configurations with a probability $\exp(-E/kT)$ and weight them evenly.

...with samples generated by the “Metropolis” algorithm

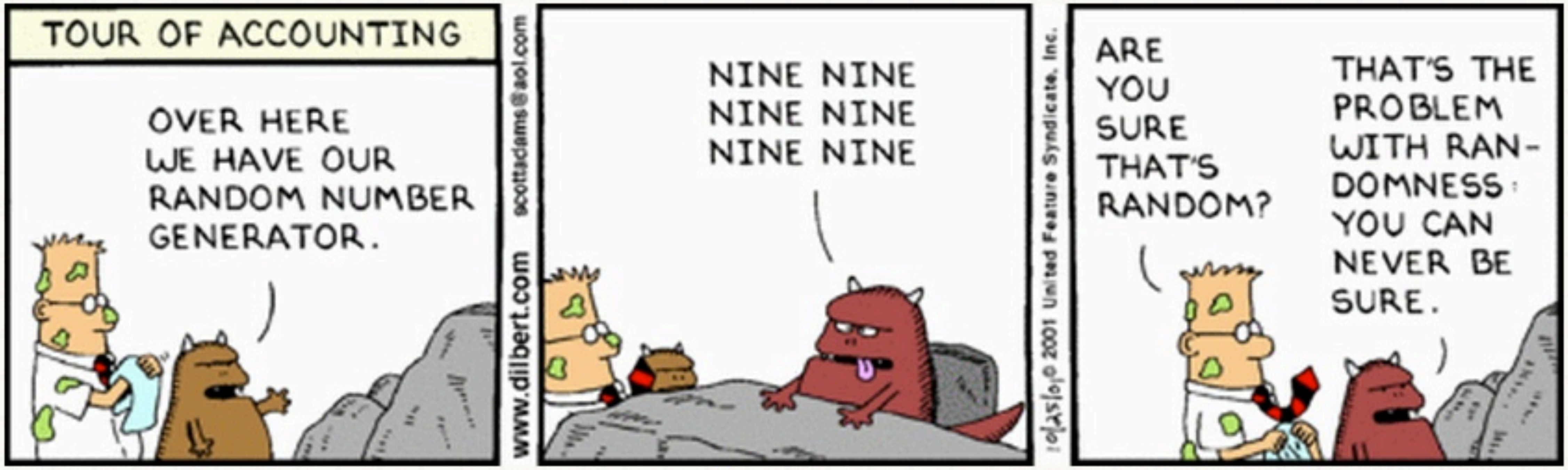


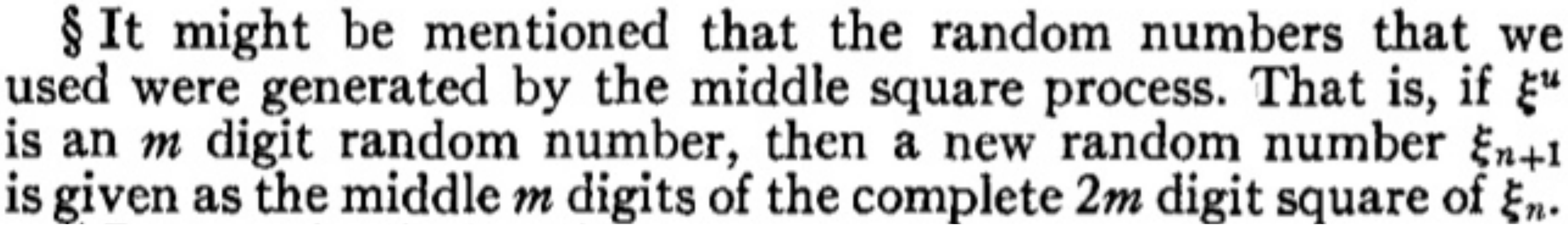
- given $x^{(r)}$, propose $x^{(r+1)}$ by random walk
- accept it if $\Delta E = E(x^{(r+1)}) - E(x^{(r)}) < 0$, or with probability $e^{-\Delta E/kT}$ if $\Delta E > 0$
- if not accepted, set $x^{(r+1)} = x^{(r)}$
- the resulting detailed balance guarantees convergence to P



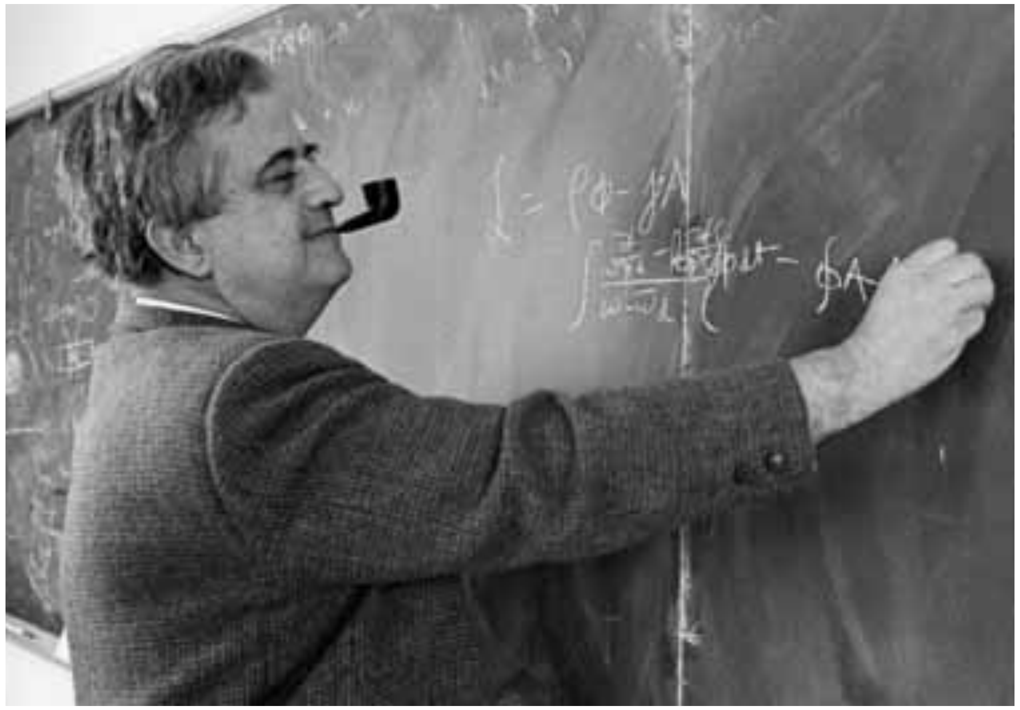
We then calculate the change in energy of the system ΔE, which is caused by the move. If $\Delta E<0$, i.e., if the move would bring the system to a state of lower energy, we allow the move and put the particle in its new position. If $\Delta E>0$, we allow the move with probability $\exp(-\Delta E/kT)$; i.e., we take a random number ξ_3 between 0 and 1, and if $\xi_3<\exp(-\Delta E/kT)$, we move the particle to its new position. If $\xi_3 >\exp(-\Delta E/kT)$, we return it to its old position.

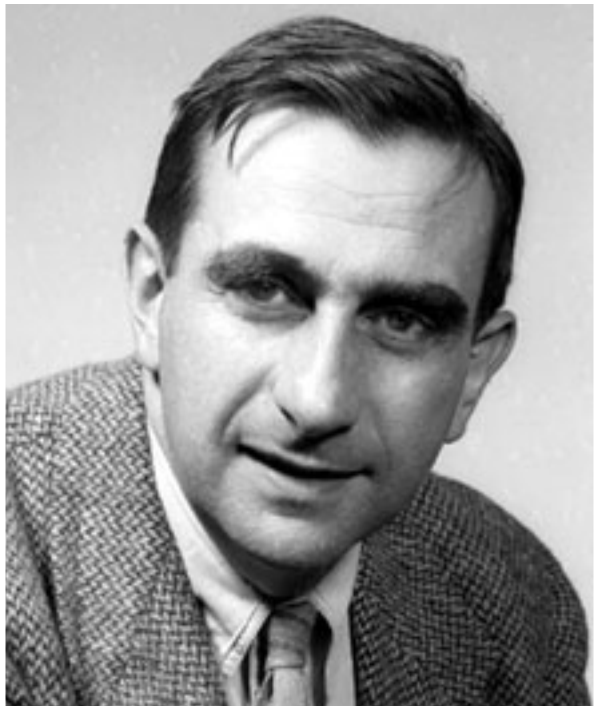
§ It might be mentioned that the random numbers that we used were generated by the middle square process. That is, if ξ^u is an m digit random number, then a new random number ξ_{n+1} is given as the middle m digits of the complete $2m$ digit square of ξ_n.


TOUR OF ACCOUNTING
OVER HERE WE HAVE OUR RANDOM NUMBER GENERATOR.
scottadams@aol.com
www.dilbert.com
NINE NINE NINE NINE NINE NINE
10/25/01 © 2001 United Feature Syndicate, Inc.
ARE YOU SURE THAT'S RANDOM?
THAT'S THE PROBLEM WITH RANDOMNESS: YOU CAN NEVER BE SURE.

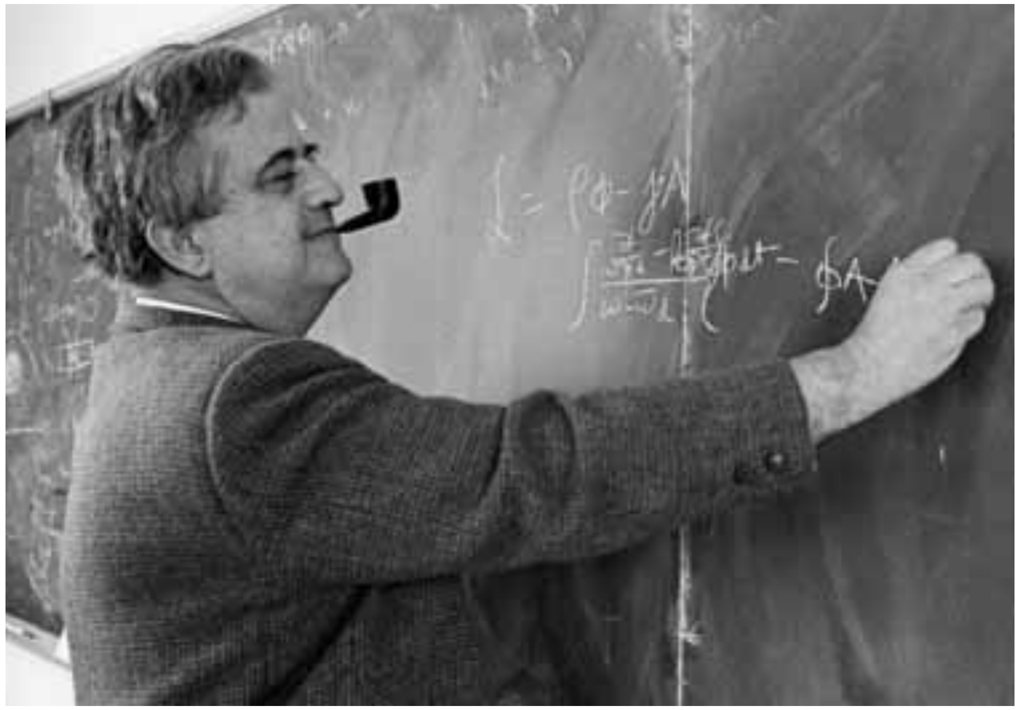


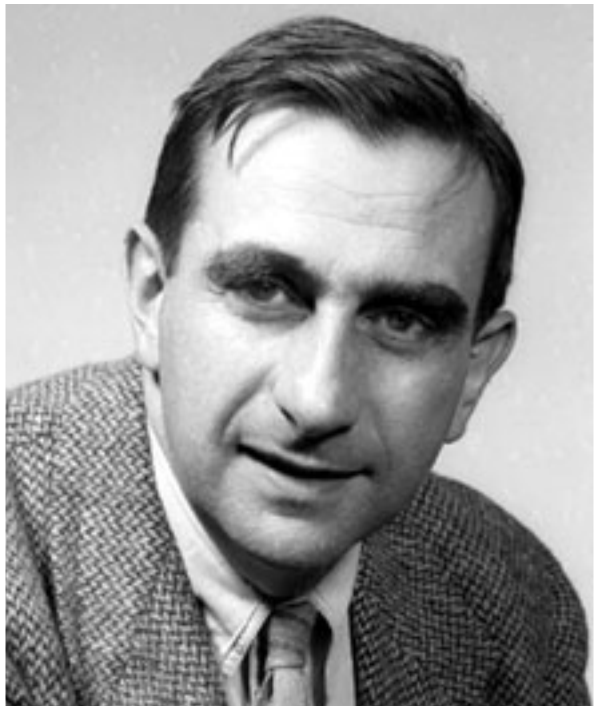
§ It might be mentioned that the random numbers that we used were generated by the middle square process. That is, if ξ^u is an m digit random number, then a new random number ξ_{n+1} is given as the middle m digits of the complete $2m$ digit square of ξ_n.





the rest is history





the rest is history

- (Metropolis–Hastings) algorithm for any *P:*
- given $x^{(r)}$, propose $x^{(r+1)}$ by $Q(x^{\text{next}};x^{\text{prev}})$
- accept it if
  $r = [P(x^{(r+1)})/P(x^{(r)})]\cdot[Q(x^{(r)};x^{(r+1)})/Q(x^{(r+1)};x^{(r)})] > 1$,
  or with probability r if $r < 1$
- if not accepted, set $x^{(r+1)} = x^{(r)}$

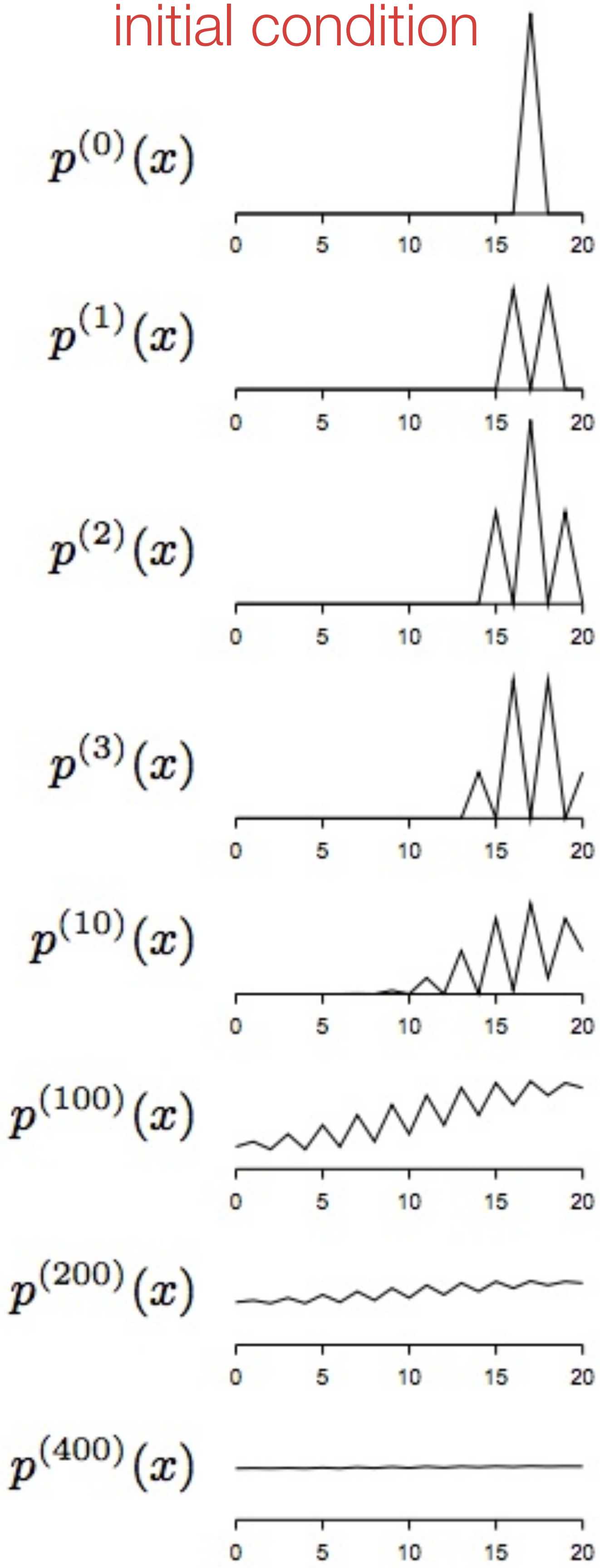
but why does it work?

- the Metropolis algorithm implements a Markov Chain $\{x^{(r)}\}$ with transition probability $T(x_i;x_j) = T_{ij}$

- the Metropolis algorithm implements a Markov Chain $\{x^{(r)}\}$ with transition probability $T(x_i;x_j) = T_{ij}$
- T is set by the proposal distribution Q and the transition rule (e.g., Metropolis)

- the Metropolis algorithm implements a Markov Chain $\{x^{(r)}\}$ with transition probability $T(x_i;x_j) = T_{ij}$
- T is set by the proposal distribution Q and the transition rule (e.g., Metropolis)
- if T_{ij} satisfies certain properties, its repeated application to any initial probability distribution ρ^0_j eventually yields the equilibrium distribution $\rho^*_i = P_i$

initial condition



equilibrium distribution

(MacKay 2003)

- if T_{ij} satisfies certain properties, its repeated application to any initial probability distribution $\rho^0{}_j$ eventually yields the equilibrium distribution $\rho^*{}_i = P_i$

- if T_{ij} satisfies certain properties, its repeated application to any initial probability distribution $\rho^0{}_j$ eventually yields the equilibrium distribution $\rho^*{}_i = P_i$

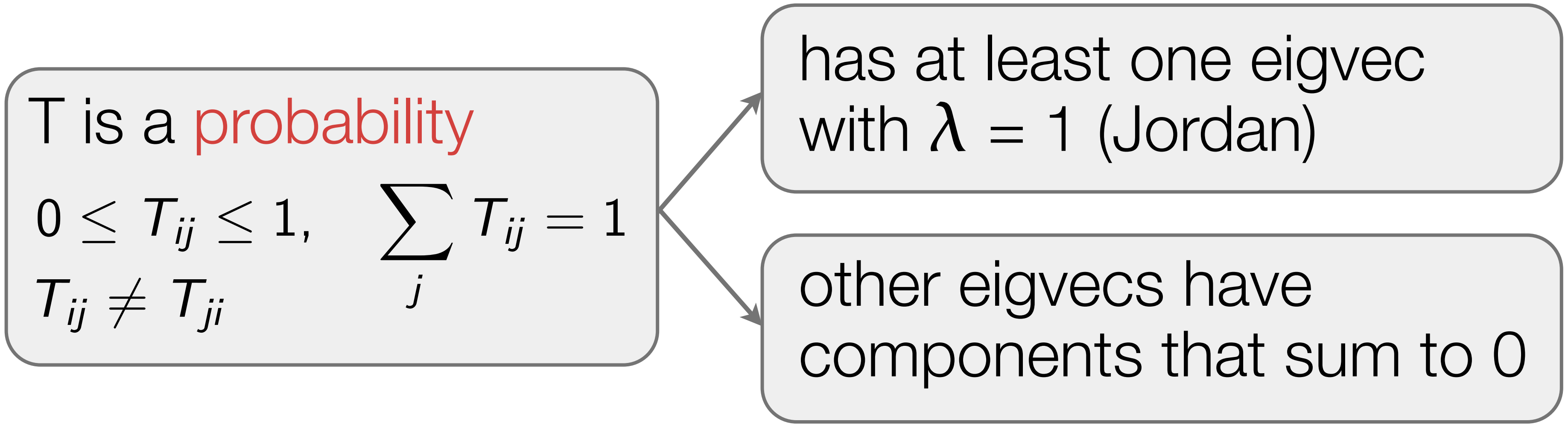
- if T_{ij} satisfies certain properties, its repeated application to any initial probability distribution $\rho^0{}_j$ eventually yields the equilibrium distribution $\rho^*{}_i = P_i$

T is a probability

$$0 \leq T_{ij} \leq 1, \quad \sum_j T_{ij} = 1$$

$$T_{ij} \neq T_{ji}$$

- if T_{ij} satisfies certain properties, its repeated application to any initial probability distribution ρ^0_j eventually yields the equilibrium distribution $\rho^*_i = P_i$



T is a probability

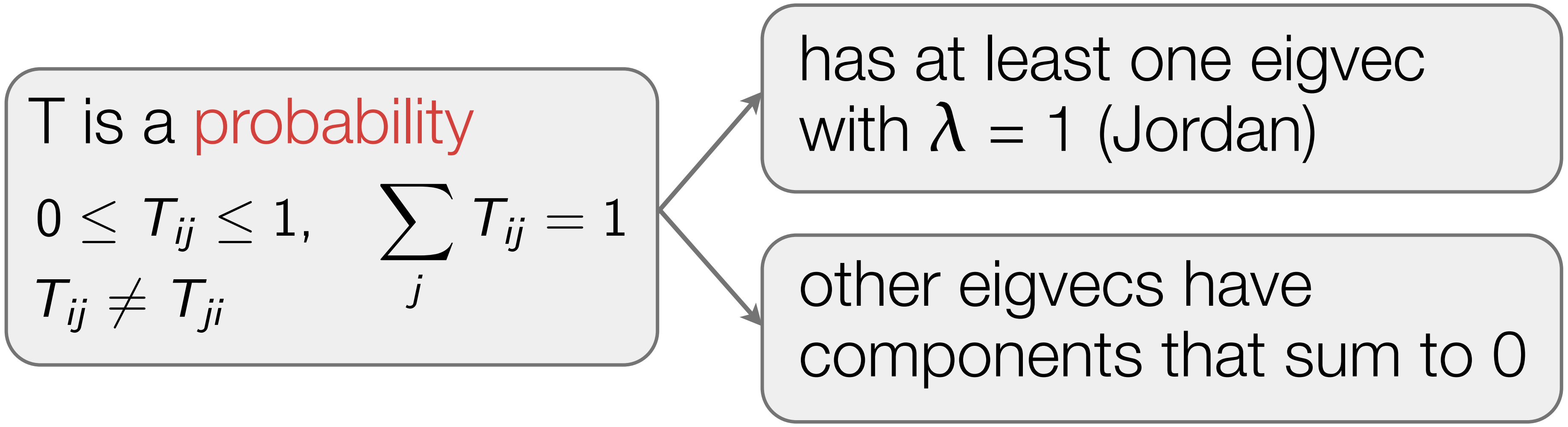
$$0 \leq T_{ij} \leq 1, \quad \sum_j T_{ij} = 1$$

$$T_{ij} \neq T_{ji}$$

has at least one eigvec with λ = 1 (Jordan)

other eigvecs have components that sum to 0

- if T_{ij} satisfies certain properties, its repeated application to any initial probability distribution $\rho^0{}_j$ eventually yields the equilibrium distribution $\rho^*{}_i = P_i$



T is a probability

$0 \leq T_{ij} \leq 1, \quad \sum_j T_{ij} = 1$

$T_{ij} \neq T_{ji}$

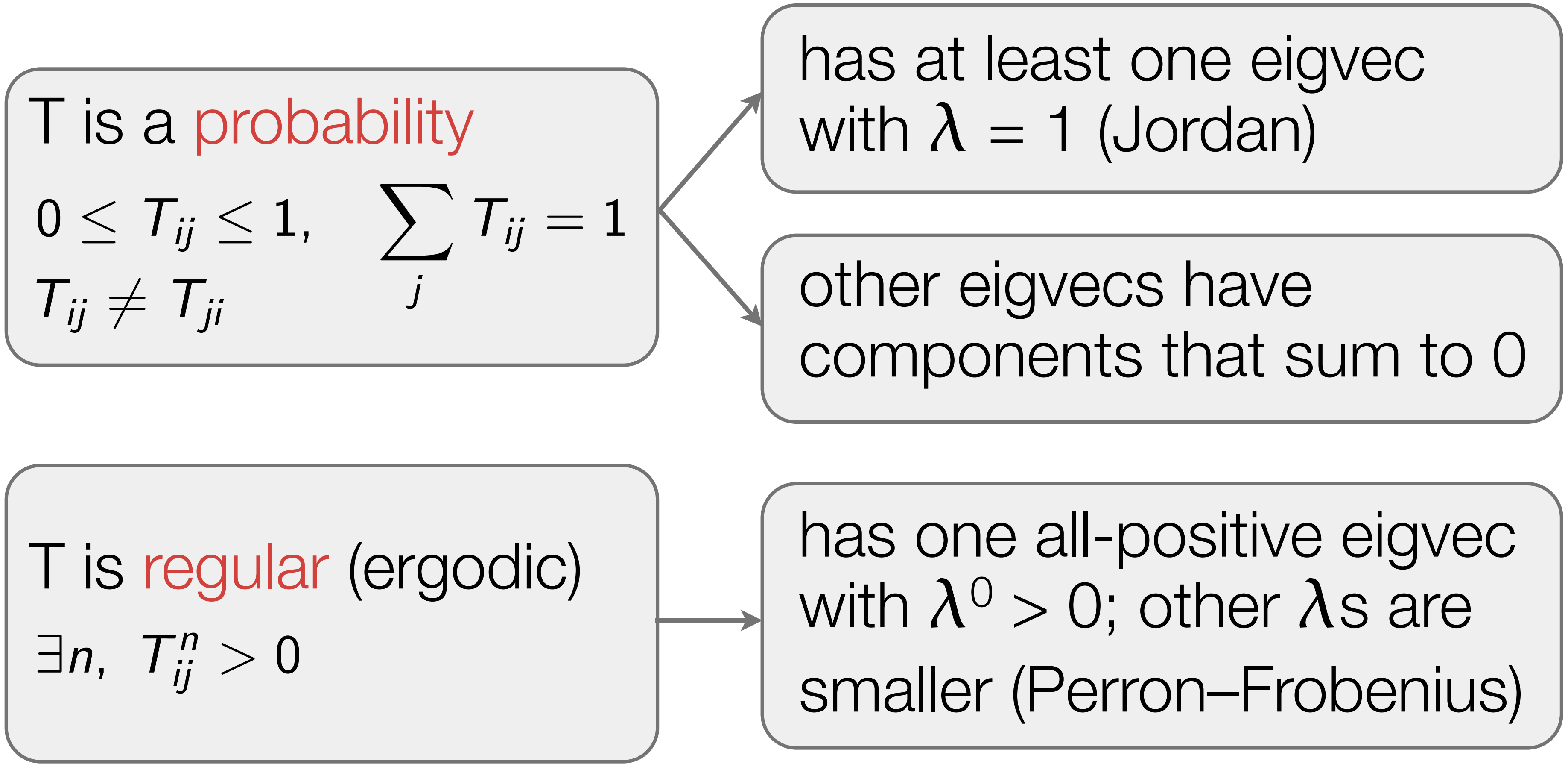
has at least one eigvec with λ = 1 (Jordan)

other eigvecs have components that sum to 0

T is regular (ergodic)

$\exists n, \;\; T^n_{ij} > 0$

- if T_{ij} satisfies certain properties, its repeated application to any initial probability distribution ρ^0_j eventually yields the equilibrium distribution $\rho^*_i = P_i$



T is a probability

$$0 \le T_{ij} \le 1, \quad \sum_j T_{ij} = 1$$

$$T_{ij} \neq T_{ji}$$

has at least one eigvec with $\lambda = 1$ (Jordan)

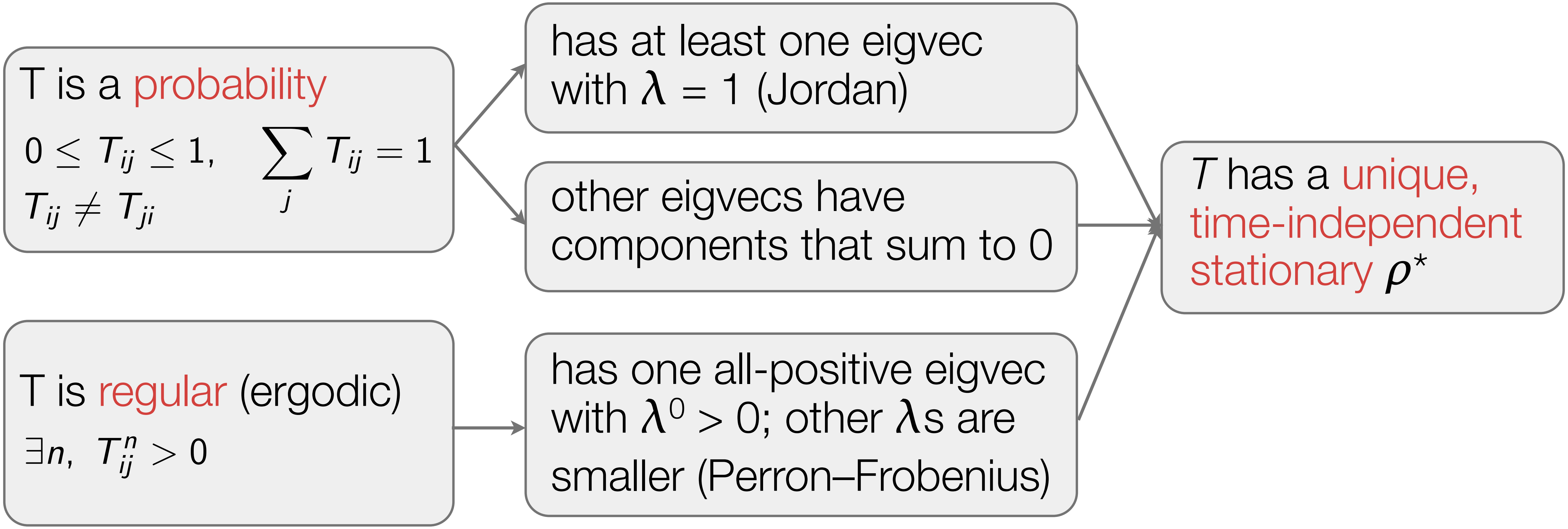
other eigvecs have components that sum to 0

T is regular (ergodic)

$$\exists n, \ T^n_{ij} > 0$$

has one all-positive eigvec with $\lambda^0 > 0$; other λs are smaller (Perron–Frobenius)

- if T_{ij} satisfies certain properties, its repeated application to any initial probability distribution $\rho^0{}_j$ eventually yields the equilibrium distribution $\rho^*{}_i = P_i$



T is a probability

$0 \leq T_{ij} \leq 1, \quad \sum_j T_{ij} = 1$

$T_{ij} \neq T_{ji}$

has at least one eigvec with λ = 1 (Jordan)

other eigvecs have components that sum to 0

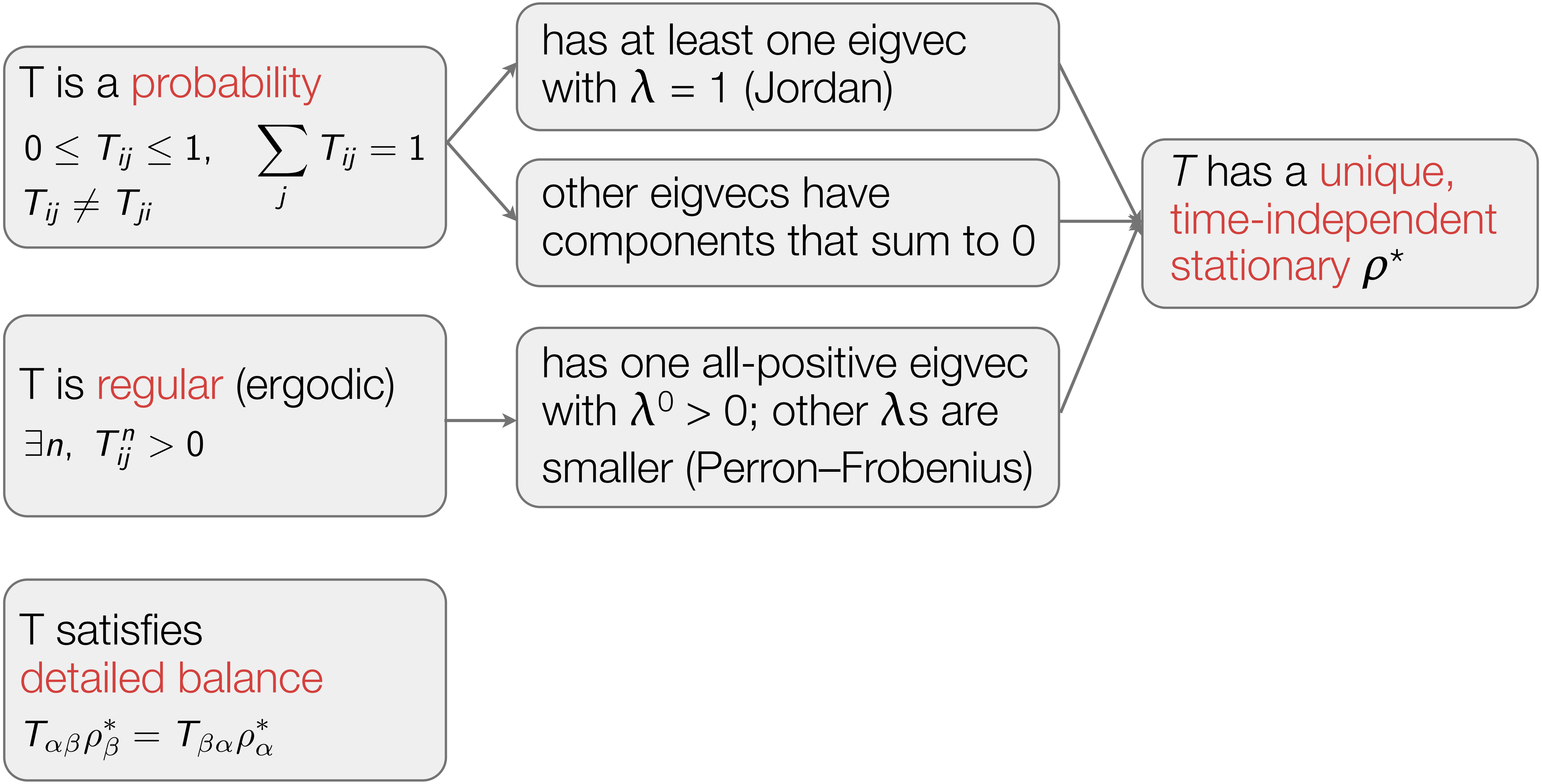
T is regular (ergodic)

$\exists n, \;\; T^n_{ij} > 0$

has one all-positive eigvec with $\lambda^0 > 0$; other λs are smaller (Perron–Frobenius)

T has a unique, time-independent stationary ρ^*

- if T_{ij} satisfies certain properties, its repeated application to any initial probability distribution ρ^0_j eventually yields the equilibrium distribution $\rho^*_i = P_i$



T is a probability

$0 \leq T_{ij} \leq 1, \quad \sum_j T_{ij} = 1$

$T_{ij} \neq T_{ji}$

has at least one eigvec with $\lambda = 1$ (Jordan)

other eigvecs have components that sum to 0

T is regular (ergodic)

$\exists n, \; T^n_{ij} > 0$

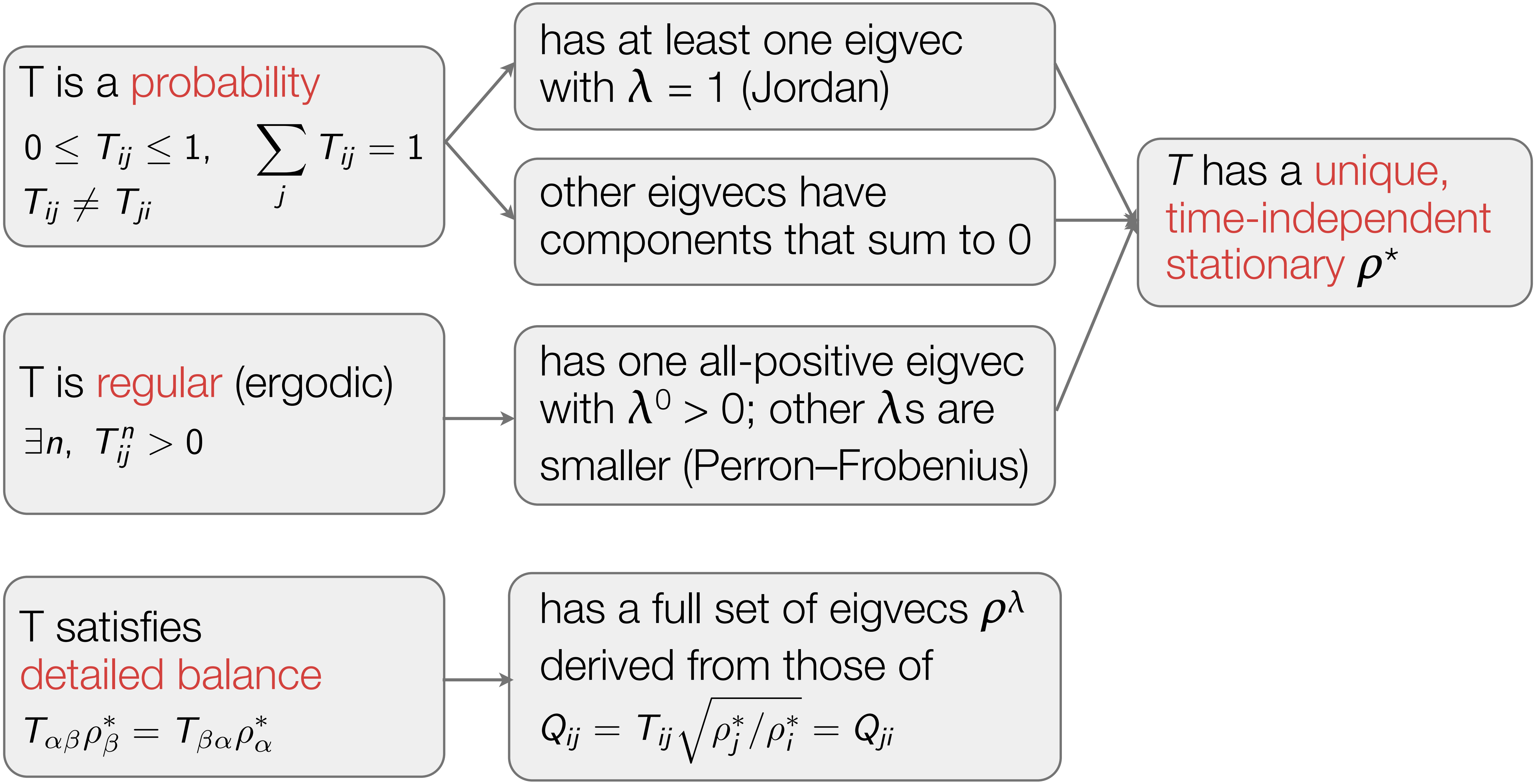
has one all-positive eigvec with $\lambda^0 > 0$; other λs are smaller (Perron–Frobenius)

T has a unique, time-independent stationary ρ^*

T satisfies detailed balance

$T_{\alpha\beta}\rho^*_\beta = T_{\beta\alpha}\rho^*_\alpha$

- if T_{ij} satisfies certain properties, its repeated application to any initial probability distribution ρ^0_j eventually yields the equilibrium distribution $\rho^*_i = P_i$



T is a probability

$0 \le T_{ij} \le 1, \quad \sum_j T_{ij} = 1$

$T_{ij} \neq T_{ji}$

has at least one eigvec with $\lambda = 1$ (Jordan)

other eigvecs have components that sum to 0

T has a unique, time-independent stationary ρ^*

T is regular (ergodic)

$\exists n, \; T^n_{ij} > 0$

has one all-positive eigvec with $\lambda^0 > 0$; other λs are smaller (Perron–Frobenius)

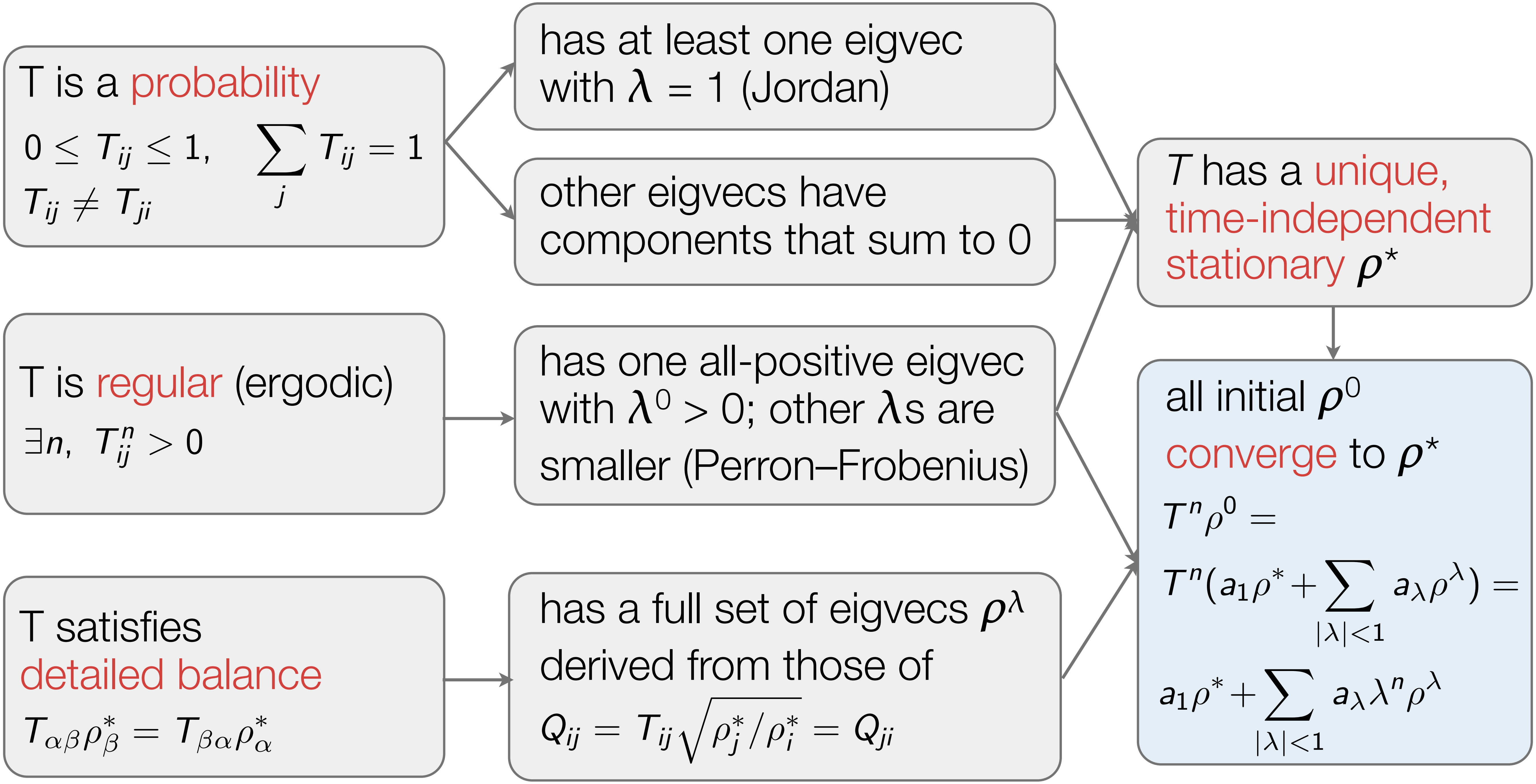
T satisfies detailed balance

$T_{\alpha\beta}\rho^*_\beta = T_{\beta\alpha}\rho^*_\alpha$

has a full set of eigvecs ρ^λ derived from those of

$Q_{ij} = T_{ij}\sqrt{\rho^*_j / \rho^*_i} = Q_{ji}$

- if T_{ij} satisfies certain properties, its repeated application to any initial probability distribution $\rho^0{}_j$ eventually yields the equilibrium distribution $\rho^*{}_i = P_i$



T is a probability

$0 \leq T_{ij} \leq 1, \quad \sum_j T_{ij} = 1$

$T_{ij} \neq T_{ji}$

has at least one eigvec with $\lambda = 1$ (Jordan)

other eigvecs have components that sum to 0

T is regular (ergodic)

$\exists n, \;\; T^n_{ij} > 0$

has one all-positive eigvec with $\lambda^0 > 0$; other λs are smaller (Perron–Frobenius)

T satisfies detailed balance

$T_{\alpha\beta}\rho^*_\beta = T_{\beta\alpha}\rho^*_\alpha$

has a full set of eigvecs ρ^λ derived from those of

$Q_{ij} = T_{ij}\sqrt{\rho^*_j/\rho^*_i} = Q_{ji}$

T has a unique, time-independent stationary ρ^*

all initial ρ^0 converge to ρ^*

$$T^n\rho^0 =$$

$$T^n(a_1\rho^* + \sum_{|\lambda|<1} a_\lambda \rho^\lambda) =$$

$$a_1\rho^* + \sum_{|\lambda|<1} a_\lambda \lambda^n \rho^\lambda$$

- the Metropolis algorithm is very general and very easy to implement

- the Metropolis algorithm is very general and very easy to implement

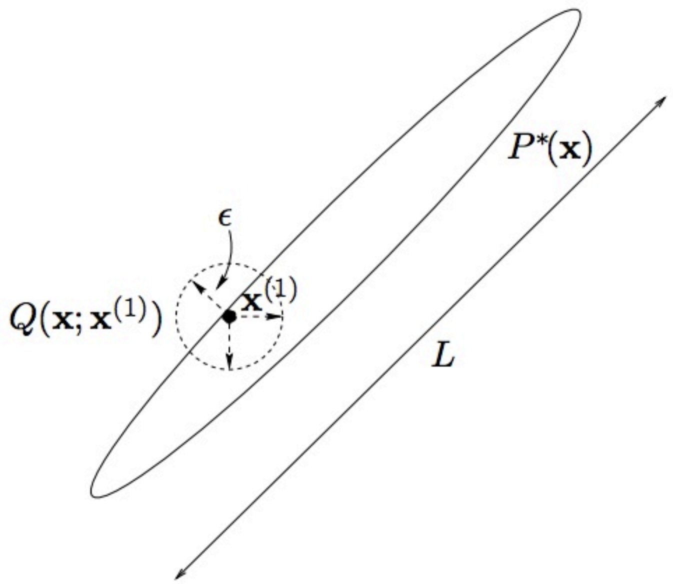
but:

- convergence, while guaranteed, is hard to assess
- random-walk exploration is very inefficient

- the Metropolis algorithm is very general and very easy to implement

but:

- convergence, while guaranteed, is hard to assess
- random-walk exploration is very inefficient



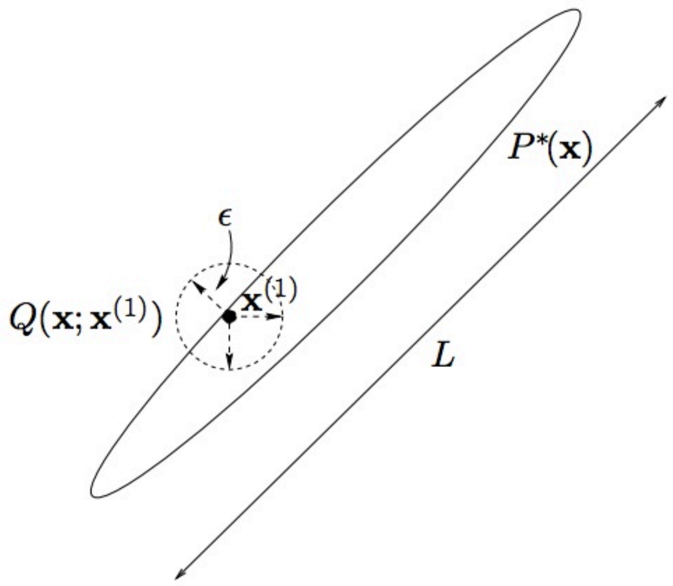
(MacKay 2003)

- need $(L/\varepsilon)^2 \sim (\sigma_{\max}/\sigma_{\min})^2$ steps to get independent sample

- the Metropolis algorithm is very general and very easy to implement

but:

- convergence, while guaranteed, is hard to assess
- random-walk exploration is very inefficient



(MacKay 2003)

- need $(L/\varepsilon)^2 \sim (\sigma_{\max}/\sigma_{\min})^2$ steps to get independent sample

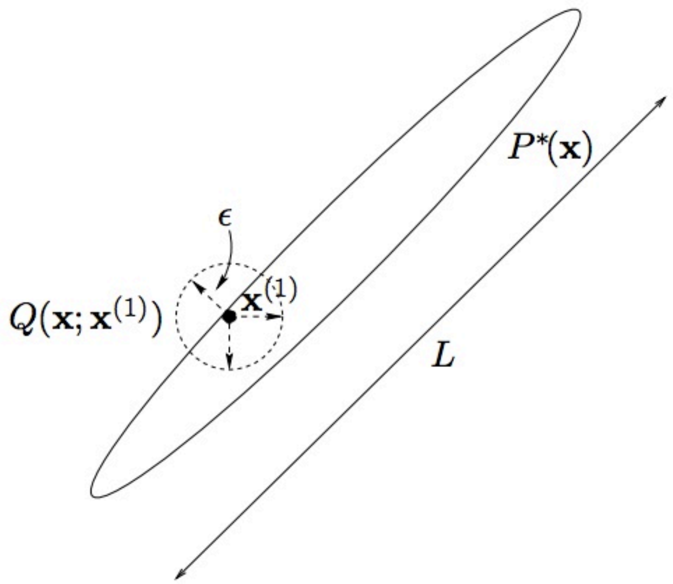
try:

- annealing, parallel tempering
- Hamiltonian MCMC
- affine-invariant samplers

- the Metropolis algorithm is very general and very easy to implement

but:

- convergence, while guaranteed, is hard to assess
- random-walk exploration is very inefficient



(MacKay 2003)

- need $(L/\varepsilon)^2 \sim (\sigma_{\text{max}}/\sigma_{\text{min}})^2$ steps to get independent sample

try:

- annealing, parallel tempering
- Hamiltonian MCMC
- affine-invariant samplers

- the evidence/partition function is difficult to compute

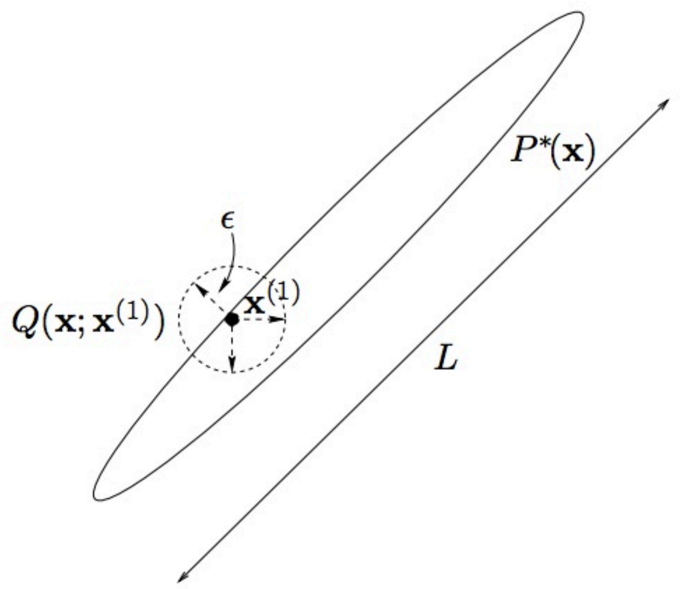
$$Z = \int e^{-E(x)/kT}\,dx,$$

$$p(M) = \int p(\text{data}|x)p(x)dx$$

- the Metropolis algorithm is very general and very easy to implement

but:

- convergence, while guaranteed, is hard to assess
- random-walk exploration is very inefficient



(MacKay 2003)

- need $(L/\varepsilon)^2 \sim (\sigma_{\max}/\sigma_{\min})^2$ steps to get independent sample

try:

- annealing, parallel tempering
- Hamiltonian MCMC
- affine-invariant samplers

- the evidence/partition function is difficult to compute

$$Z = \int e^{-E(x)/kT} dx,$$

$$p(M) = \int p(\text{data}|x)p(x)dx$$

- thermodynamic integration
- reversible-jump MCMC
- nested sampling

Nested sampling

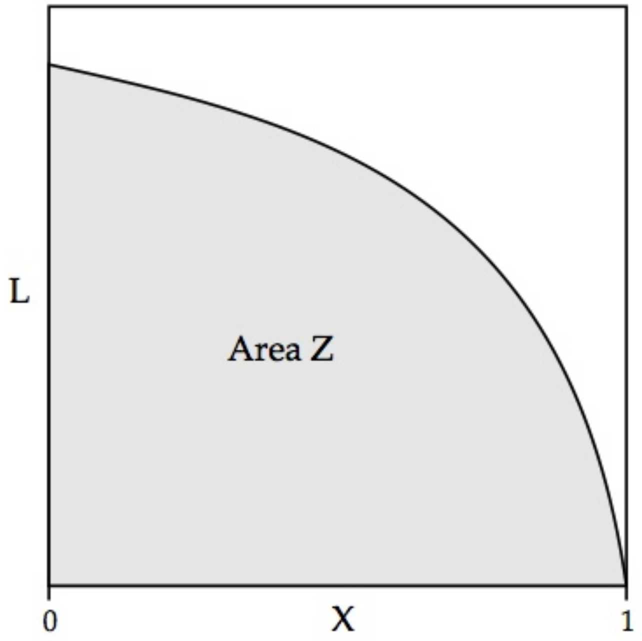
(Skilling 2006)

Nested sampling

(Skilling 2006)

$$Z = \int L(\theta)\pi(\theta)d\theta = \int L(X)dX$$

$$X(\lambda) = \int_{L(\theta)>\lambda} \pi(\theta)d\theta$$

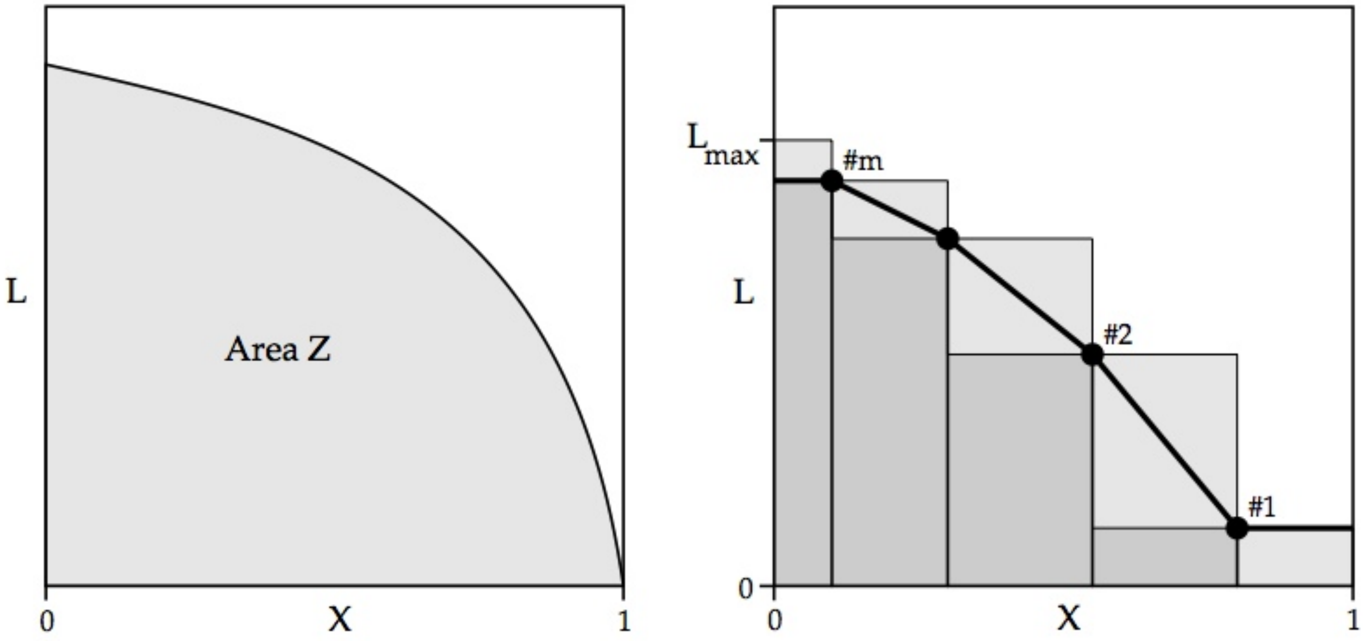


Nested sampling
(Skilling 2006)

$$Z = \int L(\theta)\pi(\theta)d\theta = \int L(X)dX$$

$$X(\lambda) = \int_{L(\theta)>\lambda} \pi(\theta)d\theta$$

$$Z = \sum_i w_i L_i$$



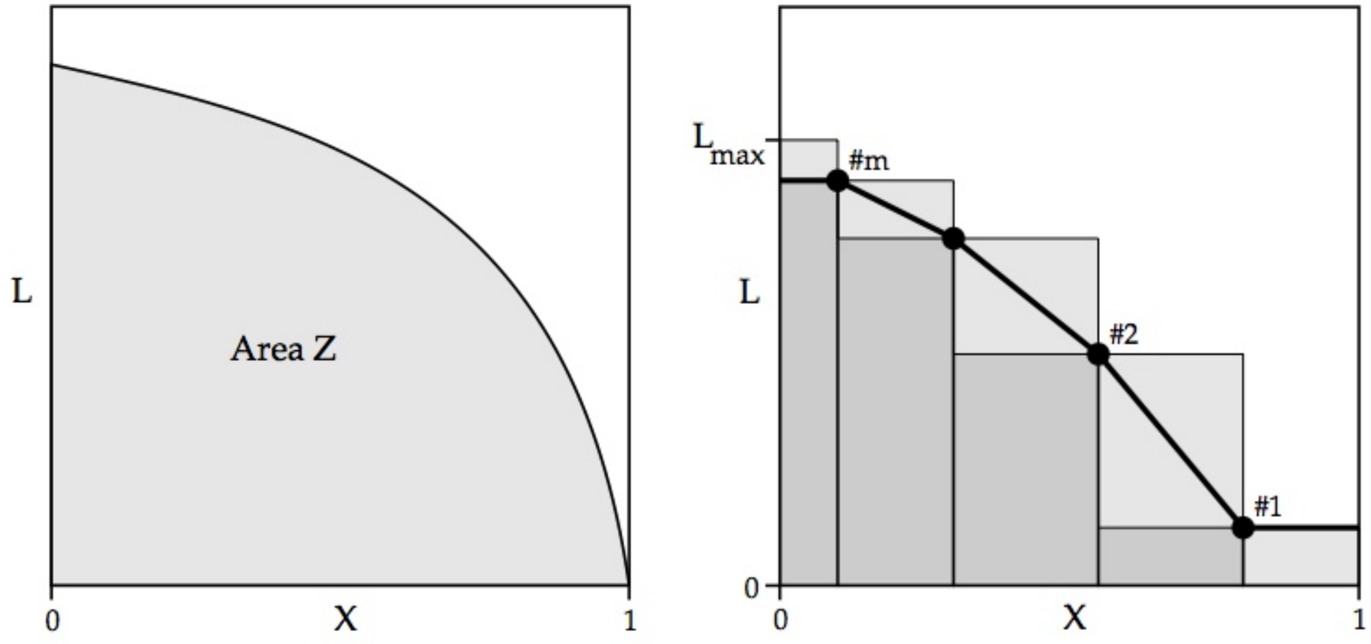
Nested sampling

(Skilling 2006)

$$Z = \int L(\theta)\pi(\theta)d\theta = \int L(X)dX$$

$$X(\lambda) = \int_{L(\theta)>\lambda} \pi(\theta)d\theta$$

$$Z = \sum_i w_i L_i$$



- *Z* is dominated by a range $-H \pm \sqrt{d}$ in log *X*, so we explore *X* linearly in log *X*

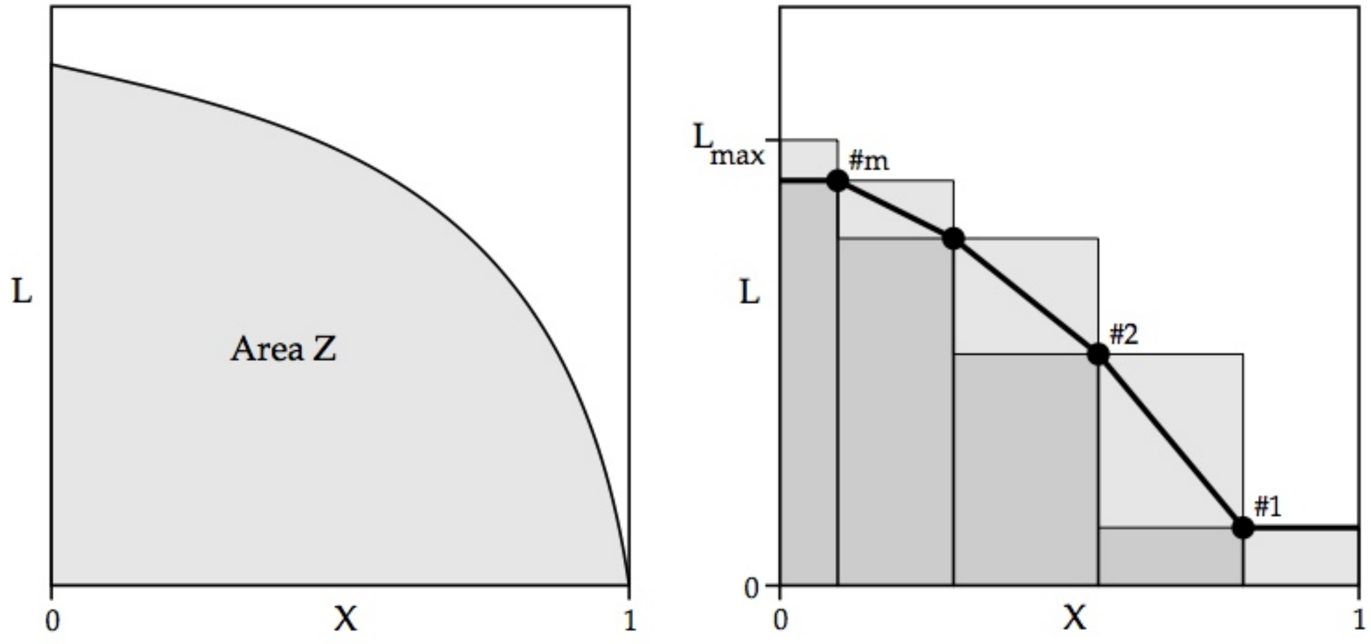
Nested sampling

(Skilling 2006)

$$Z = \int L(\theta)\pi(\theta)d\theta = \int L(X)dX$$

$$X(\lambda) = \int_{L(\theta)>\lambda} \pi(\theta)d\theta$$

$$Z = \sum_i w_i L_i$$



- *Z* is dominated by a range –*H* ± √*d* in log *X*, so we explore *X* linearly in log *X*
- choose each X_i randomly, subject to $X_i < X_{i-1}$

Nested sampling

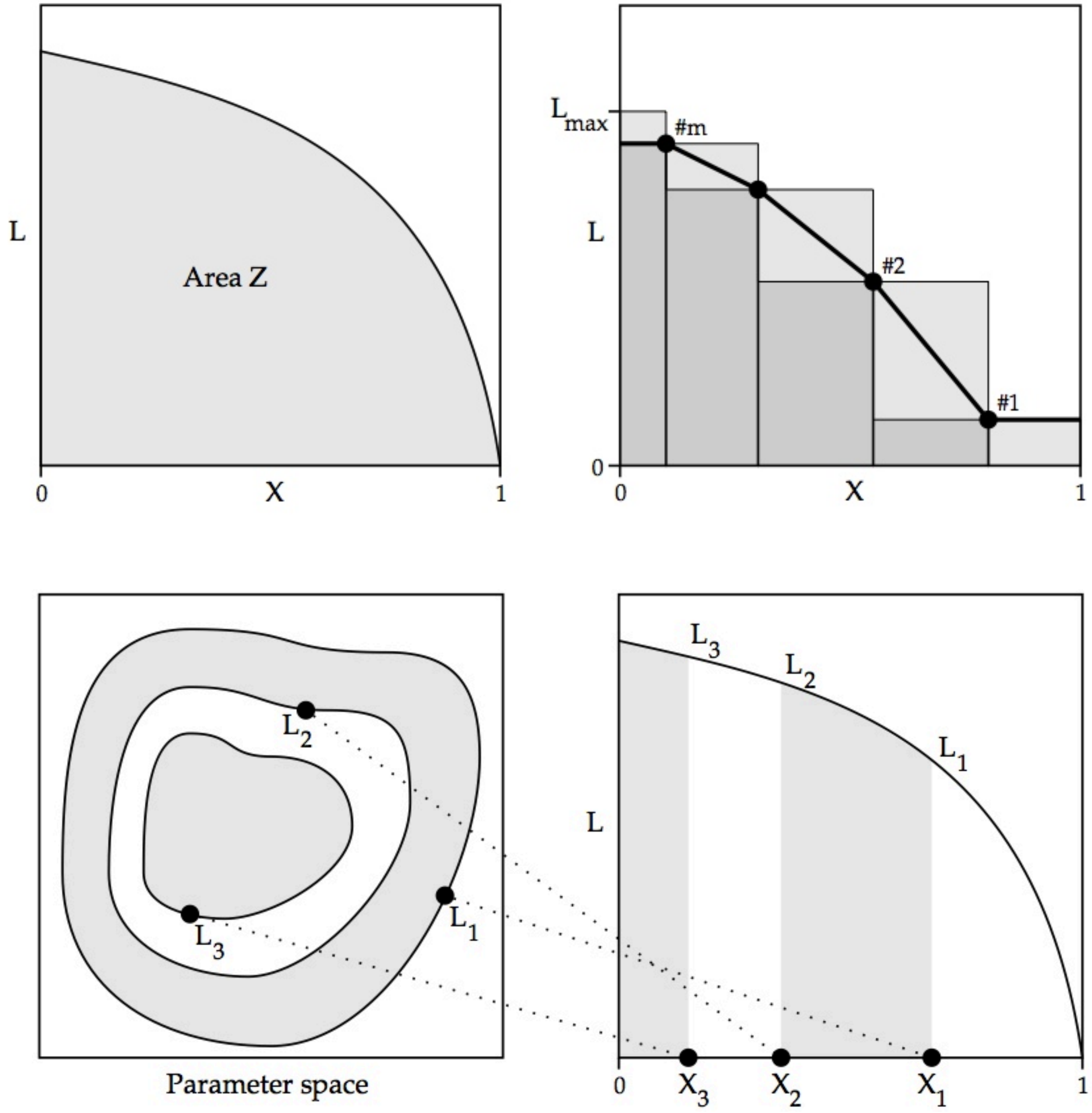
(Skilling 2006)

$$Z = \int L(\theta)\pi(\theta)d\theta = \int L(X)dX$$

$$X(\lambda) = \int_{L(\theta)>\lambda} \pi(\theta)d\theta$$

$$Z = \sum_i w_i L_i$$

- Z is dominated by a range $-H \pm \sqrt{d}$ in log X, so we explore X linearly in log X
- choose each X_i randomly, subject to $X_i < X_{i-1}$
- in practice, choose θ from $\pi(\theta)$, subject to $L(\theta_i) > L(\theta_{i-1})$



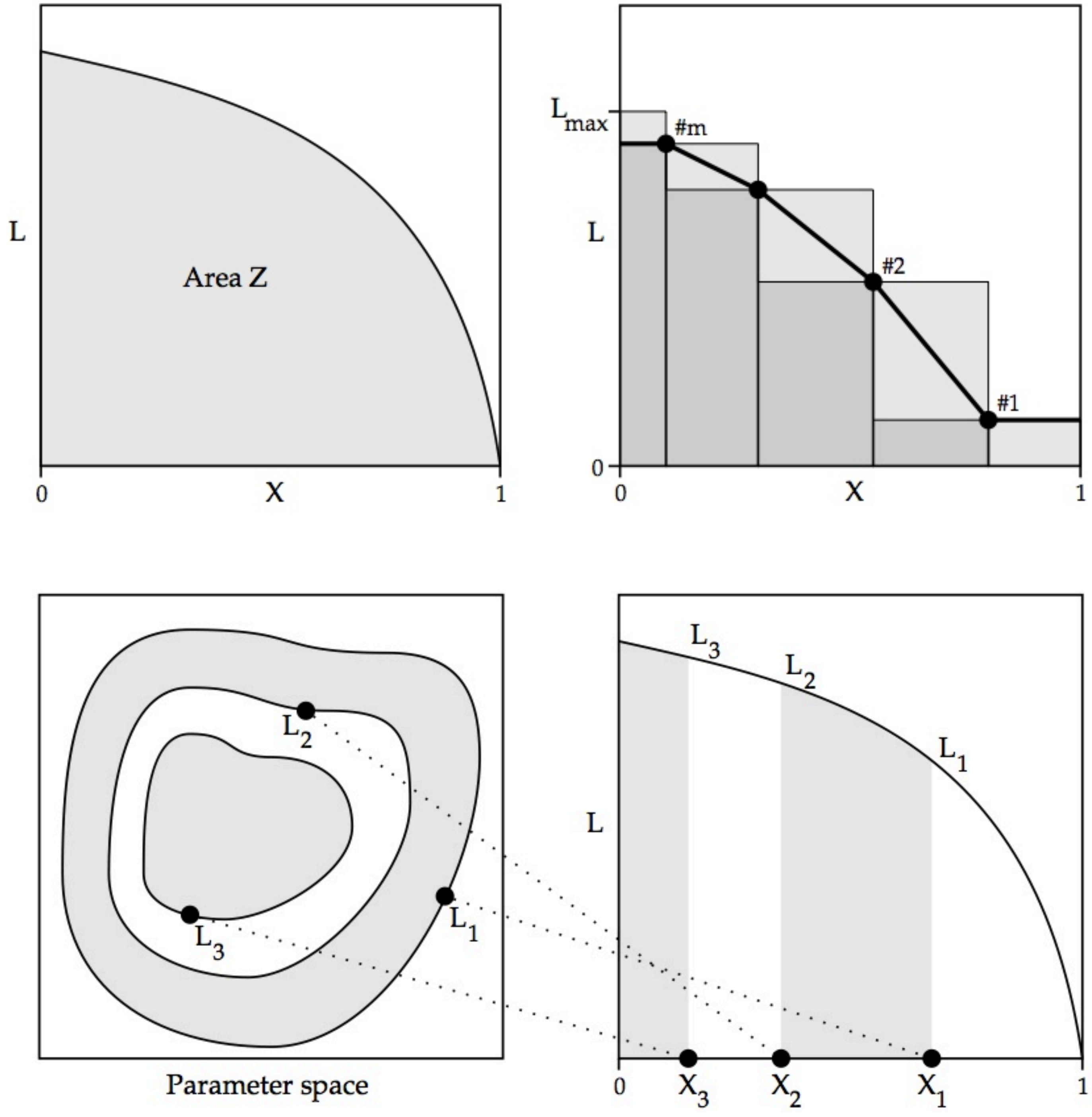
Nested sampling
(Skilling 2006)

$$Z = \int L(\theta)\pi(\theta)d\theta = \int L(X)dX$$

$$X(\lambda) = \int_{L(\theta)>\lambda} \pi(\theta)d\theta$$

$$Z = \sum_i w_i L_i$$

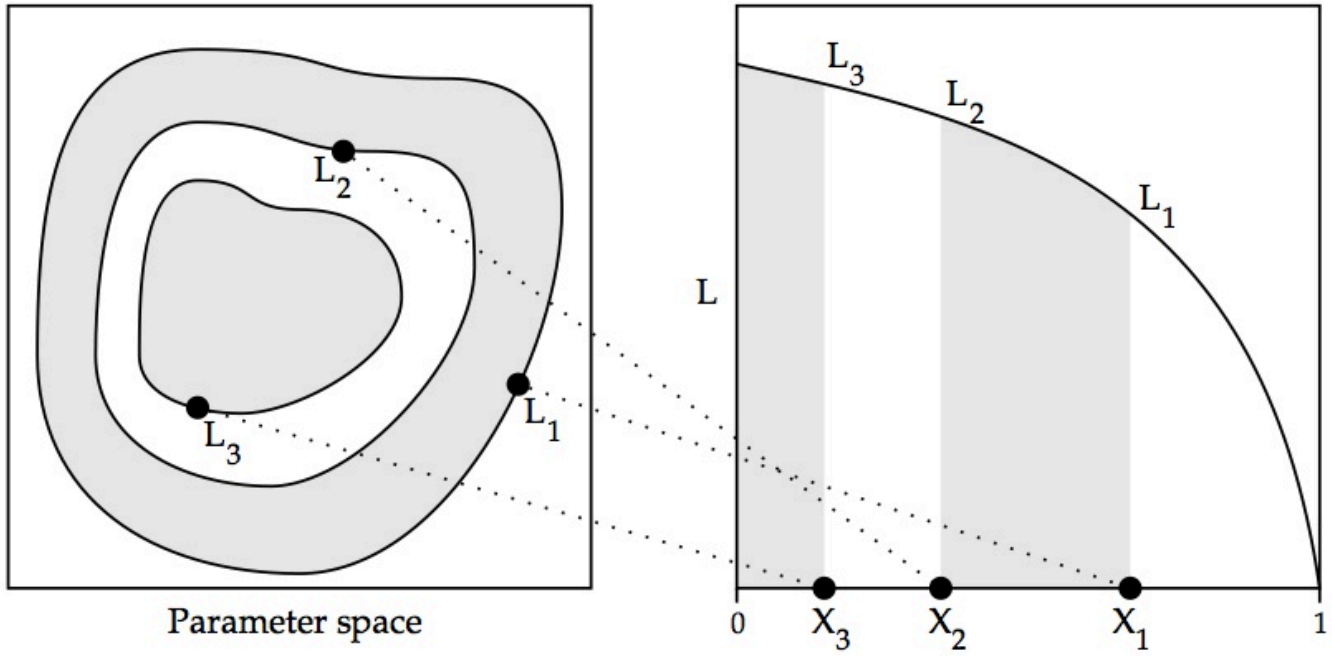
- Z is dominated by a range $-H \pm \sqrt{d}$ in log X, so we explore X linearly in log X
- choose each X_i randomly, subject to $X_i < X_{i-1}$
- in practice, choose θ from $\pi(\theta)$, subject to $L(\theta_i) > L(\theta_{i-1})$



$$X_0 = 1, \quad X_i = t_i X_{i-1}, \quad t_i \in [0,1]$$

Nested sampling
(Skilling 2006)

- Z is dominated by a range $-H \pm \sqrt{d}$ in log X, so we explore X linearly in log X
- choose each X_i randomly, subject to $X_i < X_{i-1}$
- in practice, choose θ from $\pi(\theta)$, subject to $L(\theta_i) > L(\theta_{i-1})$

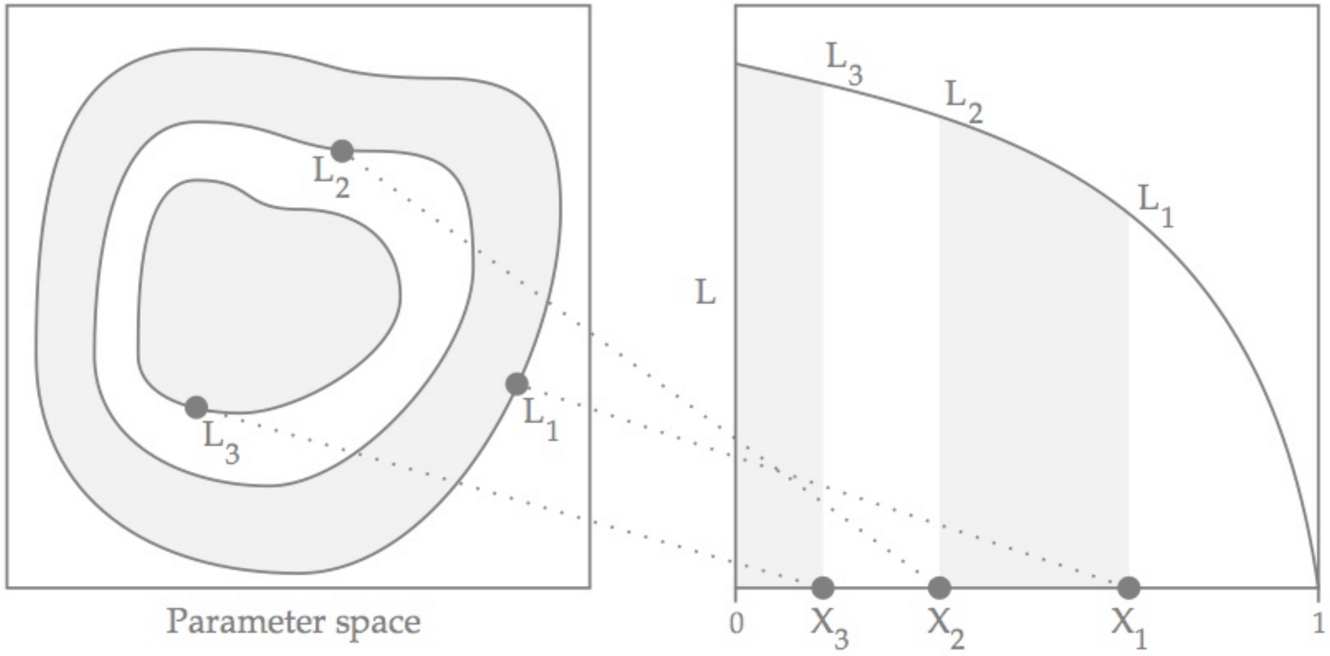


$$X_0 = 1, \quad X_i = t_i X_{i-1}, \quad t_i \in [0, 1]$$

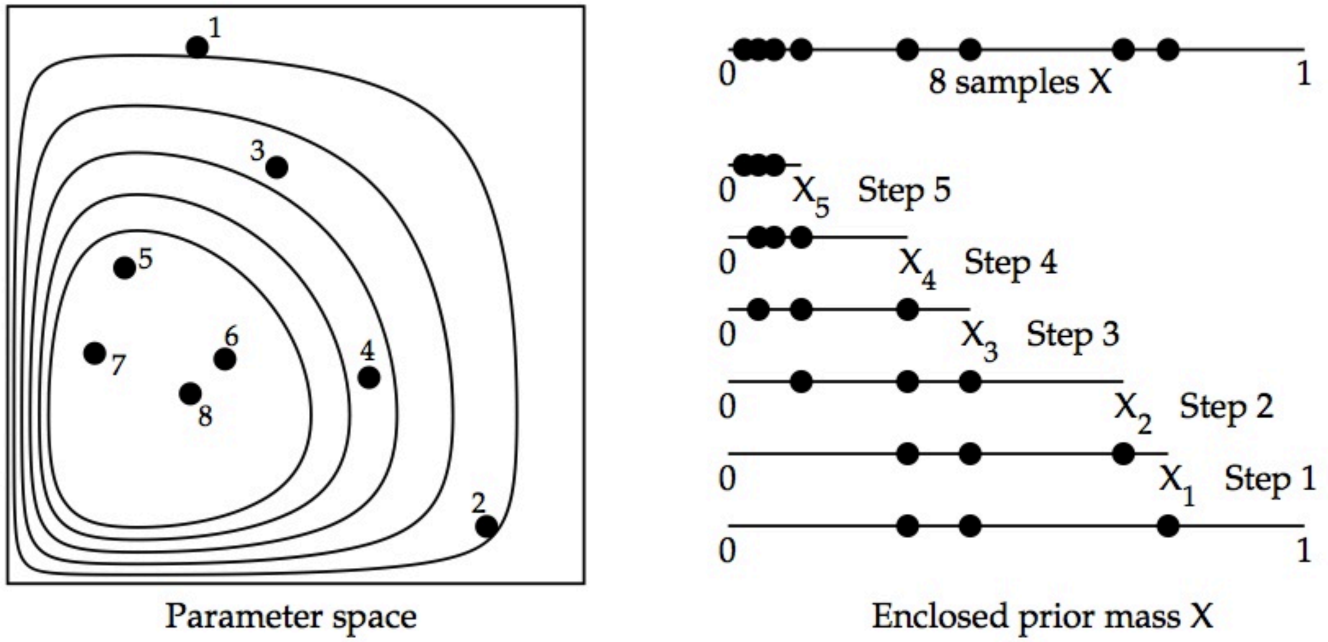
Nested sampling

(Skilling 2006)

- Z is dominated by a range $-H \pm \sqrt{d}$ in log X, so we explore X linearly in log X
- choose each X_i randomly, subject to $X_i < X_{i-1}$
- in practice, choose θ from $\pi(\theta)$, subject to $L(\theta_i) > L(\theta_{i-1})$
- working with a cloud of N live points, keep replacing the least-L member, use it as X_i



$$X_0 = 1, \quad X_i = t_i X_{i-1}, \quad t_i \in [0,1]$$

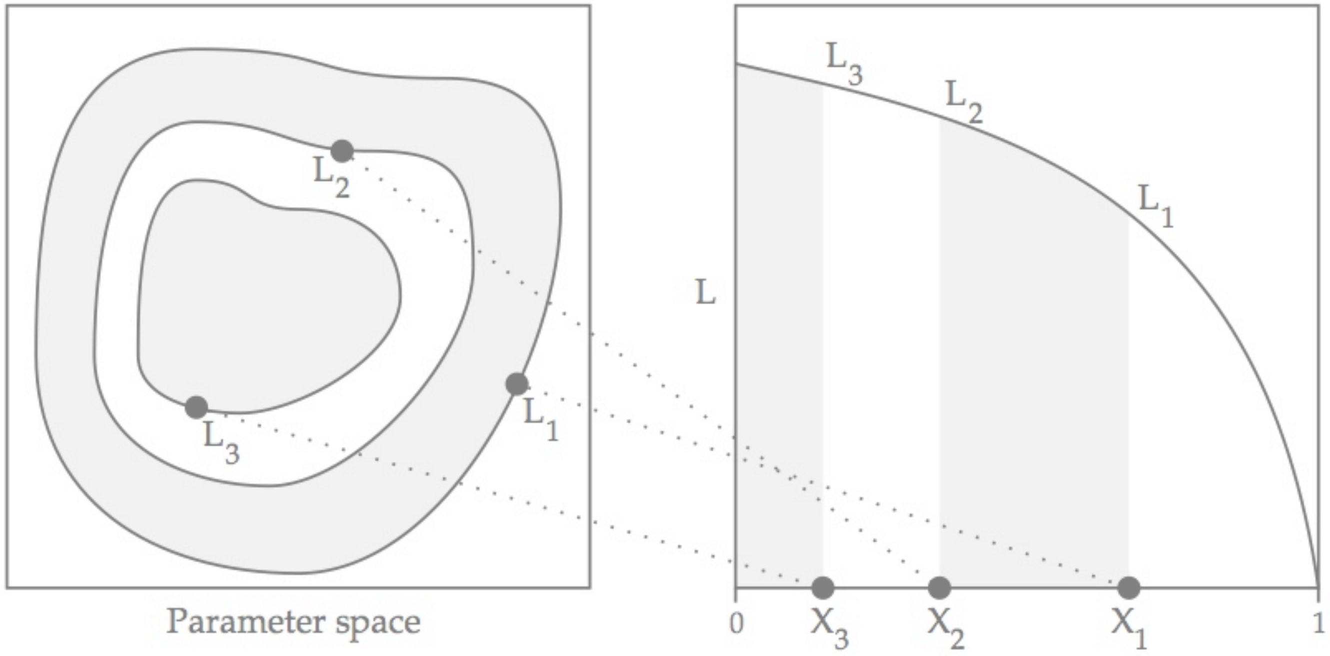


$$X_0 = 1,\ X_i = t_i X_{i-1},\ p(t_i) = N t_i^{N-1} \in [0,1]$$

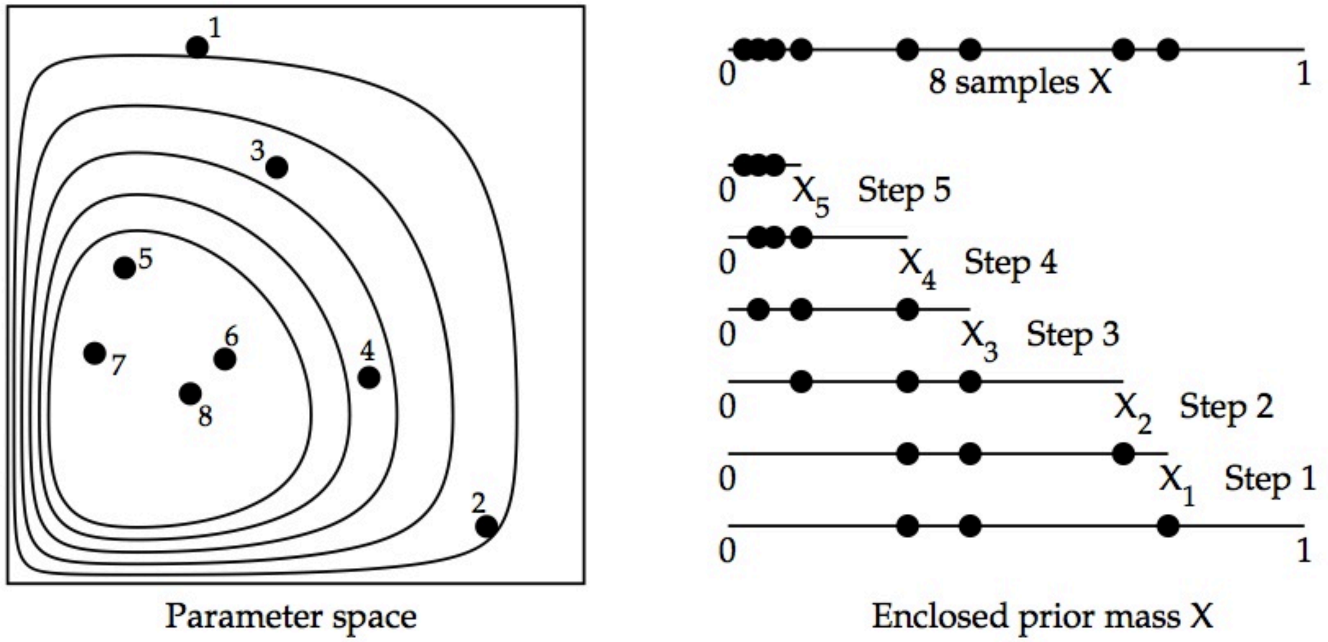
Nested sampling
(Skilling 2006)

- Z is dominated by a range $-H \pm \sqrt{d}$ in log X, so we explore X linearly in log X
- choose each X_i randomly, subject to $X_i < X_{i-1}$
- in practice, choose θ from $\pi(\theta)$, subject to $L(\theta_i) > L(\theta_{i-1})$
- working with a cloud of N live points, keep replacing the least-L member, use it as X_i
- $\log X_i \simeq (-i \pm \sqrt{i})/N$

$$Z = \sum_i w_i L_i, \;\; w_i = \frac{X_{i-1} - X_{i+1}}{2}$$



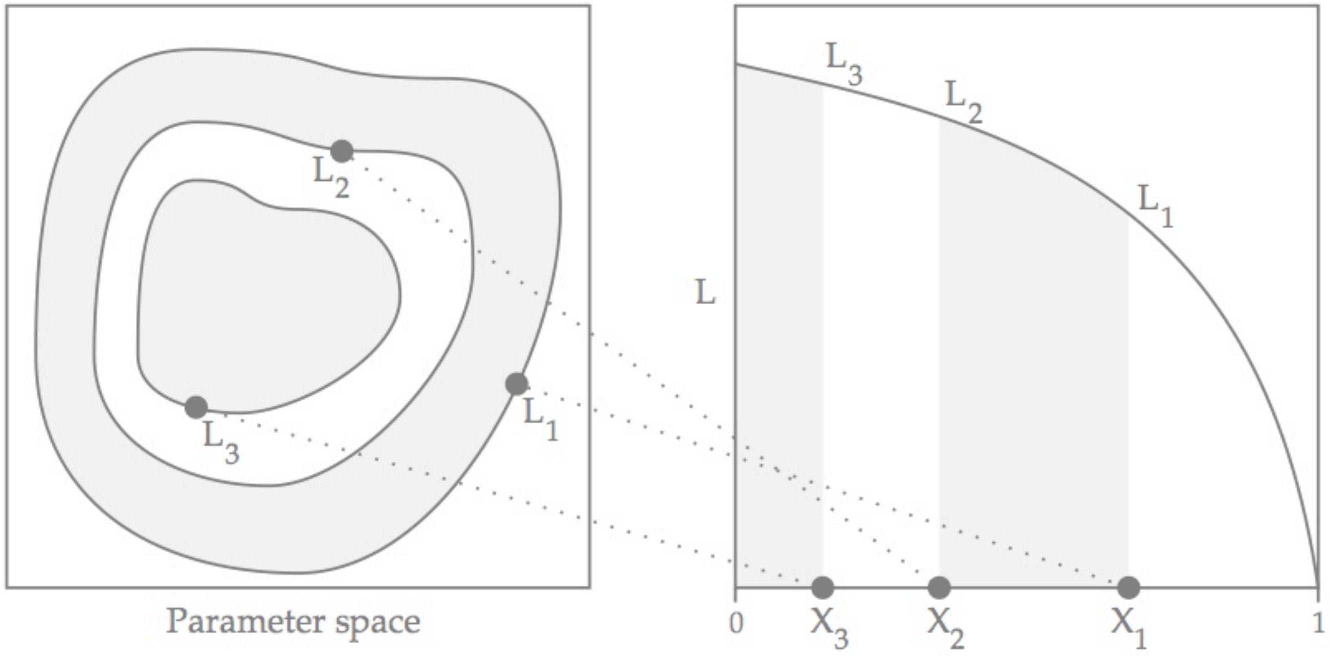
$$X_0 = 1, \quad X_i = t_i X_{i-1}, \quad t_i \in [0,1]$$



$$X_0 = 1, \; X_i = t_i X_{i-1}, \; p(t_i) = N t_i^{N-1} \in [0,1]$$

Nested sampling
(Skilling 2006)

- Z is dominated by a range $-H \pm \sqrt{d}$ in log X, so we explore X linearly in log X
- choose each X_i randomly, subject to $X_i < X_{i-1}$
- in practice, choose θ from $\pi(\theta)$, subject to $L(\theta_i) > L(\theta_{i-1})$

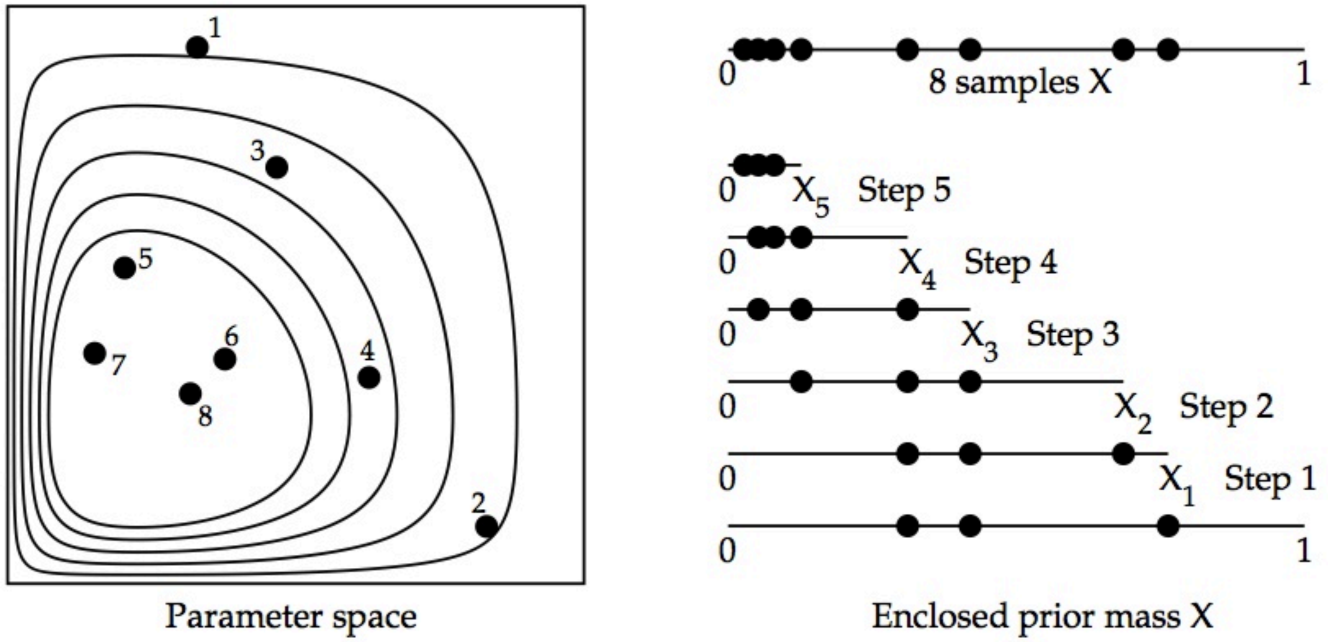


$$X_0 = 1, \quad X_i = t_i X_{i-1}, \quad t_i \in [0,1]$$

- working with a cloud of N live points, keep replacing the least-L member, use it as X_i
- $\log X_i \simeq (-i \pm \sqrt{i})/N$

$$Z = \sum_i w_i L_i, \ \ w_i = \frac{X_{i-1} - X_{i+1}}{2}$$

- stop when Z converges

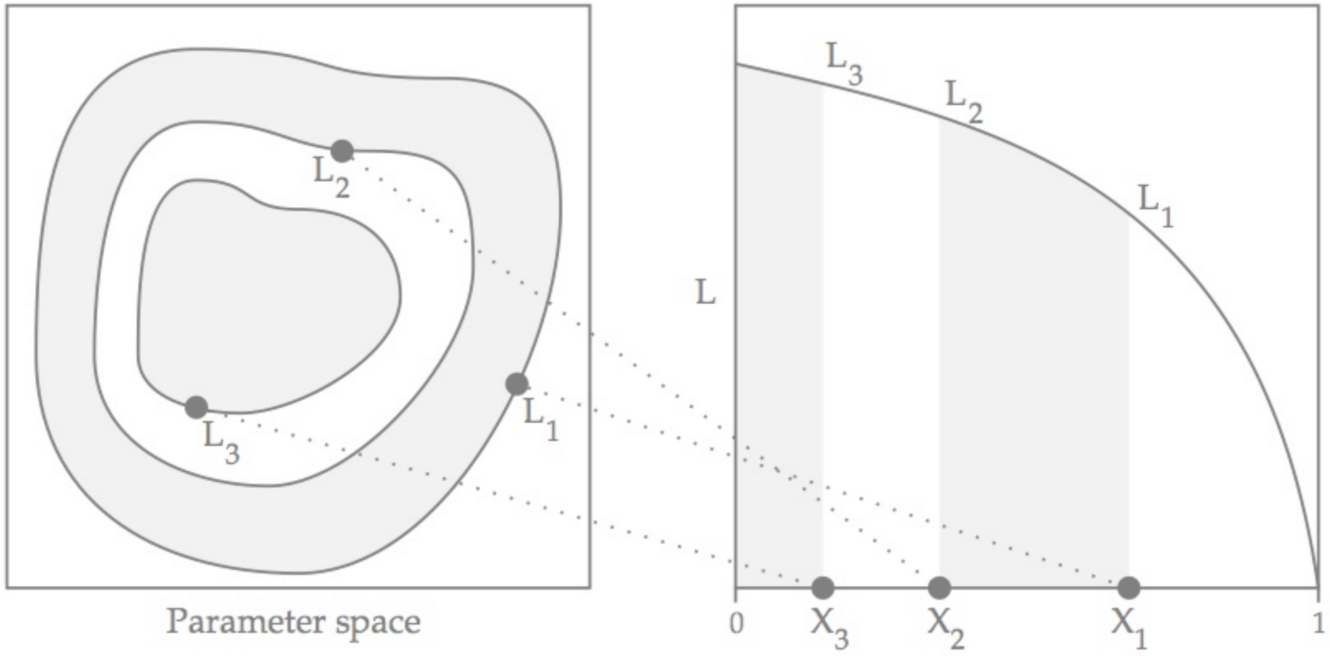


$$X_0 = 1, \ X_i = t_i X_{i-1}, \ p(t_i) = N t_i^{N-1} \in [0,1]$$

Nested sampling
(Skilling 2006)

see bit.ly/multinest
by Farhan Feroz, Hobson, Bridges

- Z is dominated by a range $-H \pm \sqrt{d}$ in $\log X$, so we explore X linearly in $\log X$
- choose each X_i randomly, subject to $X_i < X_{i-1}$
- in practice, choose θ from $\pi(\theta)$, subject to $L(\theta_i) > L(\theta_{i-1})$

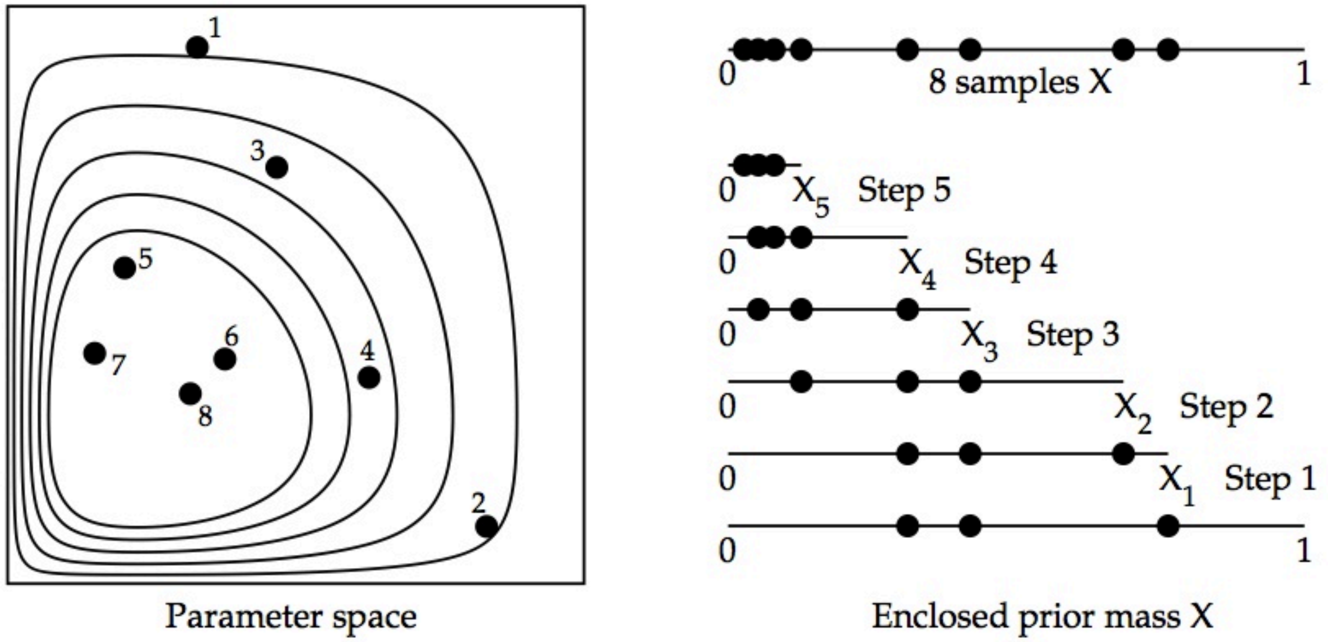


$$X_0 = 1, \quad X_i = t_i X_{i-1}, \quad t_i \in [0,1]$$

- working with a cloud of N live points, keep replacing the least-L member, use it as X_i
- $\log X_i \simeq (-i \pm \sqrt{i})/N$

$$Z = \sum_i w_i L_i, \;\; w_i = \frac{X_{i-1} - X_{i+1}}{2}$$

- stop when Z converges

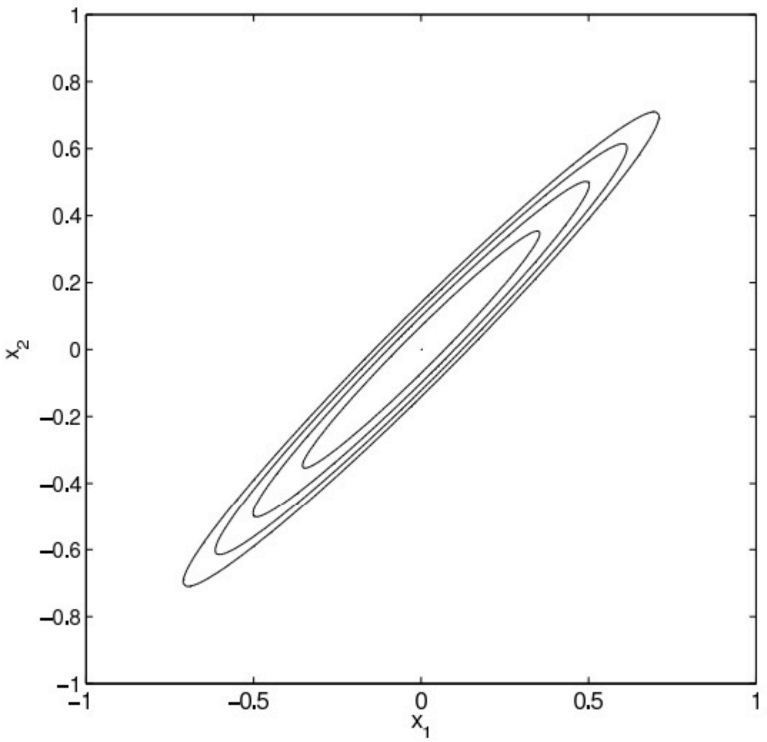


$$X_0 = 1, \; X_i = t_i X_{i-1}, \; p(t_i) = N t_i^{N-1} \in [0,1]$$

Affine–invariant sampling

(Goodman–Weare 2010)

Affine–invariant sampling
(Goodman–Weare 2010)

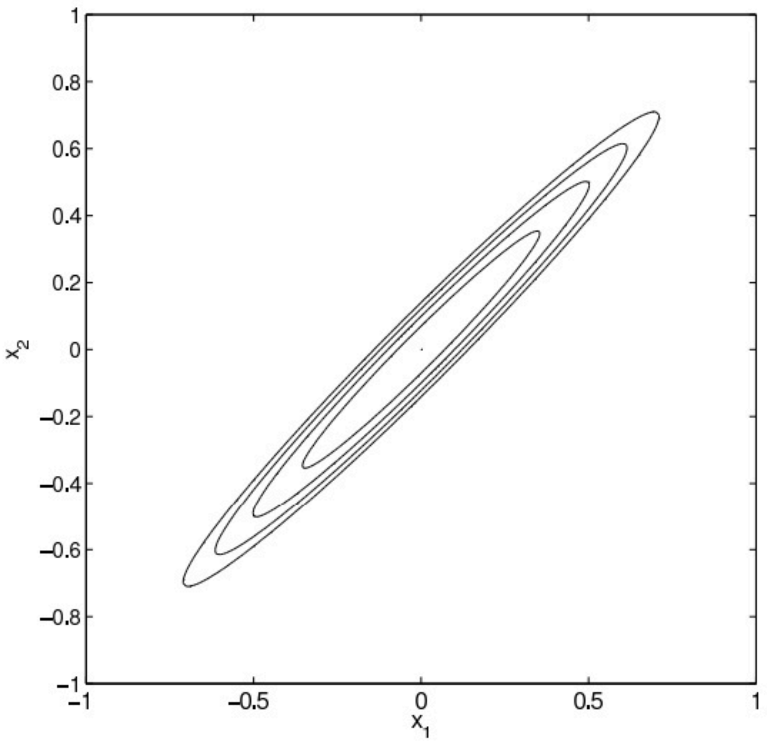


$$\pi(x) \propto \exp\left(\frac{-(x_1 - x_2)^2}{2\epsilon} - \frac{(x_1 + x_2)^2}{2}\right)$$

hard to sample!

Affine–invariant sampling

(Goodman–Weare 2010)



$$\pi(x) \propto \exp\left(\frac{-(x_1 - x_2)^2}{2\epsilon} - \frac{(x_1 + x_2)^2}{2}\right)$$

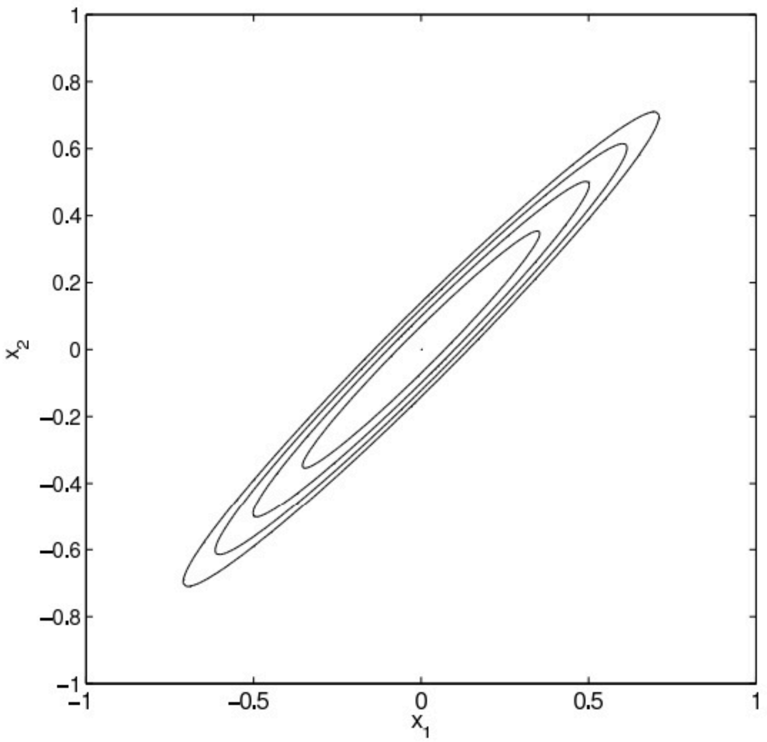
hard to sample!

$$y_1 = \frac{x_1 - x_2}{\sqrt{\epsilon}}, \qquad y_2 = x_1 + x_2$$

$$\pi_A(y) \propto \exp\left(-\frac{(y_1 + y_2)^2}{2}\right)$$

much better!

Affine–invariant sampling
(Goodman–Weare 2010)



$$\pi(x) \propto \exp\left(\frac{-(x_1 - x_2)^2}{2\epsilon} - \frac{(x_1 + x_2)^2}{2}\right)$$

hard to sample!

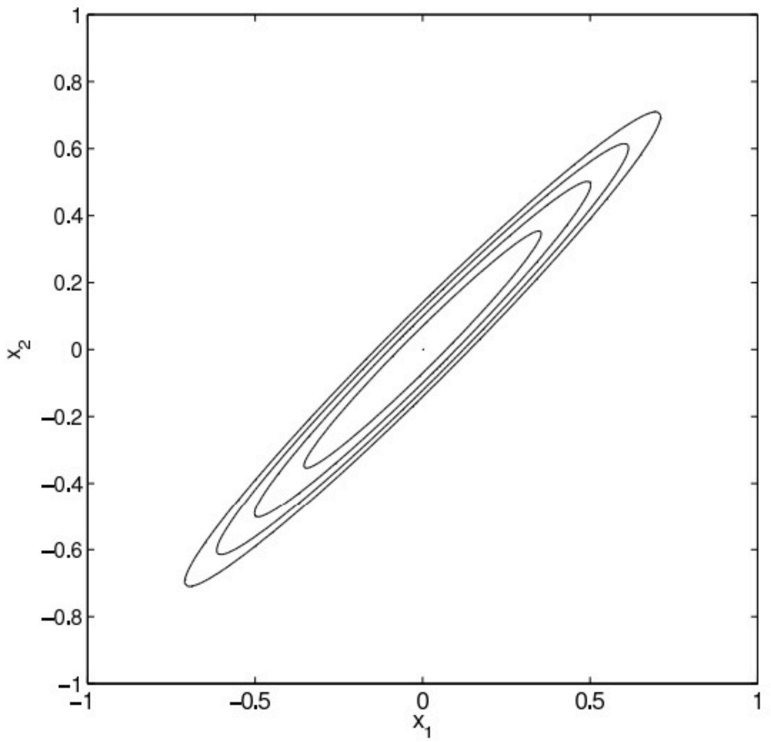
$$y_1 = \frac{x_1 - x_2}{\sqrt{\epsilon}}, \qquad y_2 = x_1 + x_2$$

$$\pi_A(y) \propto \exp\left(-\frac{(y_1 + y_2)^2}{2}\right)$$

much better!

- a sampler with affine-invariant transition probabilities would conform automatically to any coordinates (cf. simplex optimization)

Affine–invariant sampling
(Goodman–Weare 2010)



$$\pi(x) \propto \exp\left(\frac{-(x_1 - x_2)^2}{2\epsilon} - \frac{(x_1 + x_2)^2}{2}\right)$$

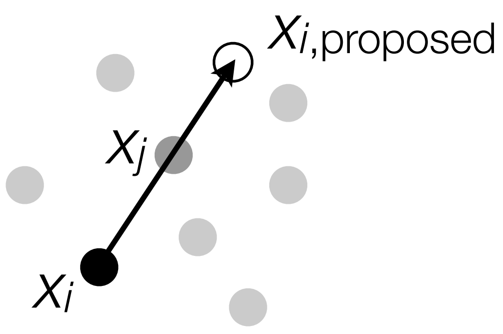
hard to sample!

$$y_1 = \frac{x_1 - x_2}{\sqrt{\epsilon}}, \qquad y_2 = x_1 + x_2$$

$$\pi_A(y) \propto \exp\left(-\frac{(y_1 + y_2)^2}{2}\right)$$

much better!

- a sampler with affine-invariant transition probabilities would conform automatically to any coordinates (cf. simplex optimization)
- G–W ensemble sampler: to update each walker x_i (i=1,..,N) in turn

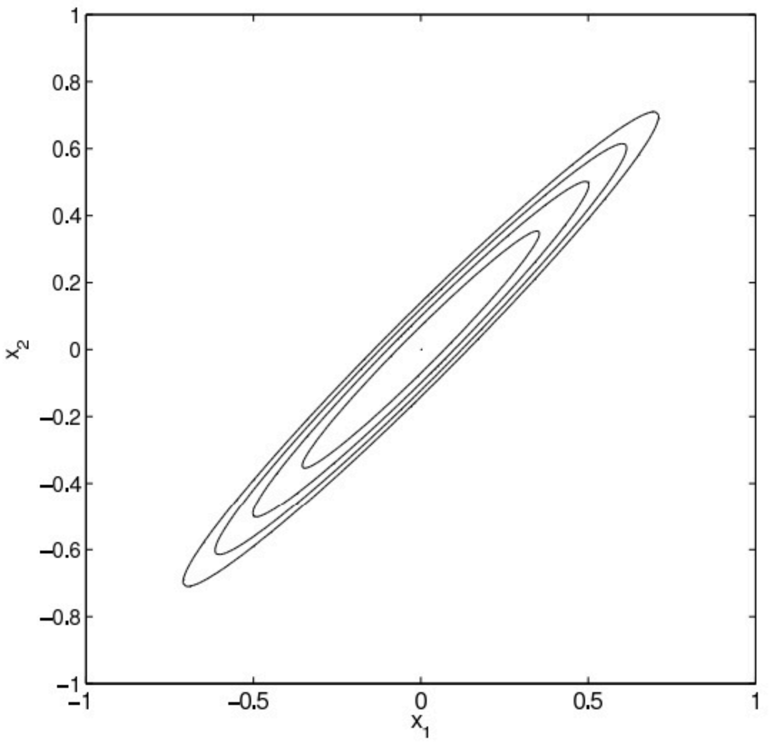


propose $x_i \longrightarrow x_j + z\cdot(x_k - x_j),\ k \neq j$

with density $g(z) \propto 1/\sqrt{z}$ for $z \in [1/a, a]$

accept with Metropolis

Affine–invariant sampling
(Goodman–Weare 2010)



$$\pi(x) \propto \exp\left(\frac{-(x_1-x_2)^2}{2\epsilon} - \frac{(x_1+x_2)^2}{2}\right)$$

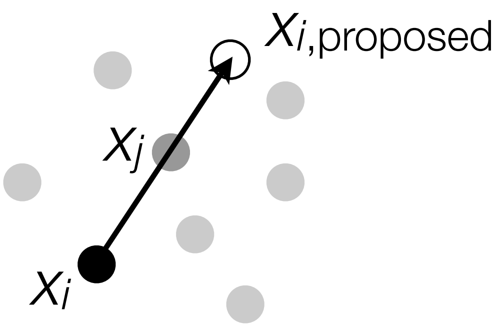
hard to sample!

$$y_1 = \frac{x_1 - x_2}{\sqrt{\epsilon}}, \qquad y_2 = x_1 + x_2$$

$$\pi_A(y) \propto \exp\left(-\frac{(y_1+y_2)^2}{2}\right)$$

much better!

- a sampler with affine-invariant transition probabilities would conform automatically to any coordinates (cf. simplex optimization)
- G–W ensemble sampler: to update each walker x_i (i=1,..,N) in turn



propose $x_i \longrightarrow x_j + z\cdot(x_k - x_j)$, $k \neq j$

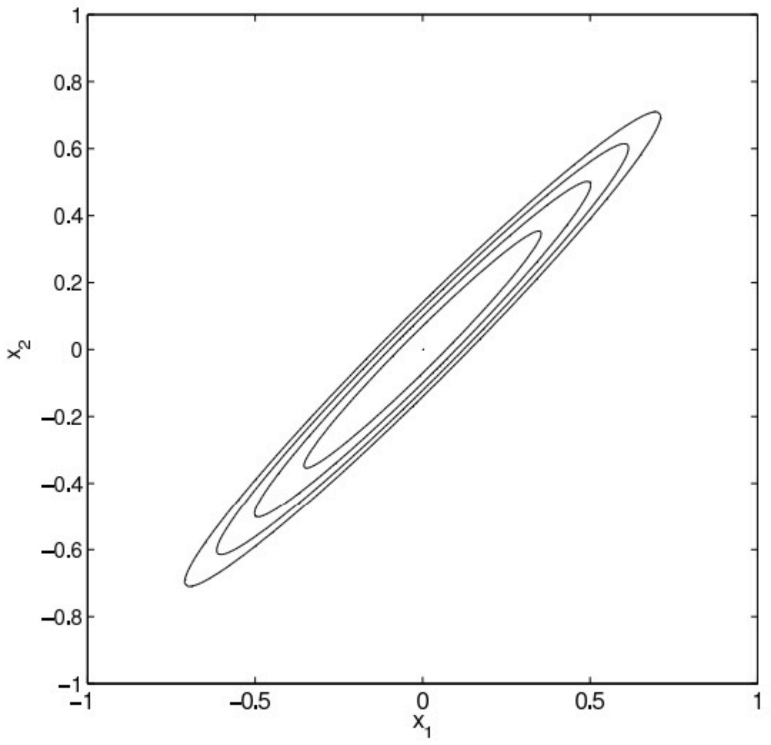
with density $g(z) \propto 1/\sqrt{z}$ for $z \in [1/a, a]$

accept with Metropolis

- the final ensemble approximates the posterior

Affine–invariant sampling
(Goodman–Weare 2010)

see dan.iel.fm/emcee
by Daniel Foreman-Mackey *et al.*



$$\pi(x) \propto \exp\left(\frac{-(x_1 - x_2)^2}{2\epsilon} - \frac{(x_1 + x_2)^2}{2}\right)$$

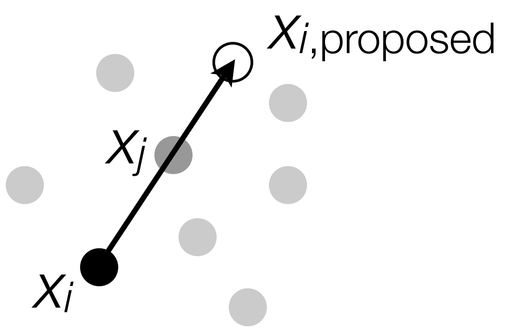
hard to sample!

$$y_1 = \frac{x_1 - x_2}{\sqrt{\epsilon}}, \qquad y_2 = x_1 + x_2$$

$$\pi_A(y) \propto \exp\left(-\frac{(y_1 + y_2)^2}{2}\right)$$

much better!

- a sampler with affine-invariant transition probabilities would conform automatically to any coordinates (cf. simplex optimization)
- G–W ensemble sampler: to update each walker x_i (i=1,..,N) in turn
  propose $x_i \longrightarrow x_j + z \cdot (x_k - x_j)$, $k \neq j$
  with density $g(z) \propto 1/\sqrt{z}$ for $z \in [1/a, a]$
  accept with Metropolis



- the final ensemble approximates the posterior

Discussion

Discussion

God is always in the details...

Discussion

God is always in the details...

...but the transcendent is experienced, never proved

Discussion

God is always in the details...

...but the transcendent is experienced, never proved

Trust no one...

Discussion

God is always in the details...

...but the transcendent is experienced, never proved

Trust no one...

...because there’s no free lunch

Discussion

God is always in the details...

...but the transcendent is experienced, never proved

Trust no one...

...because there’s no free lunch

(not even with genetically engineered algorithms)

Discussion

God is always in the details...

...but the transcendent is experienced, never proved

Trust no one...

...because there’s no free lunch

(not even with genetically engineered algorithms)

Parallelization is hard...

Discussion

God is always in the details...

...but the transcendent is experienced, never proved

Trust no one...

...because there’s no free lunch

(not even with genetically engineered algorithms)

Parallelization is hard...

...but Gaussian integrals are easy

Discussion

God is always in the details...

...but the transcendent is experienced, never proved

Trust no one...

...because there’s no free lunch

(not even with genetically engineered algorithms)

Parallelization is hard...

...but Gaussian integrals are easy

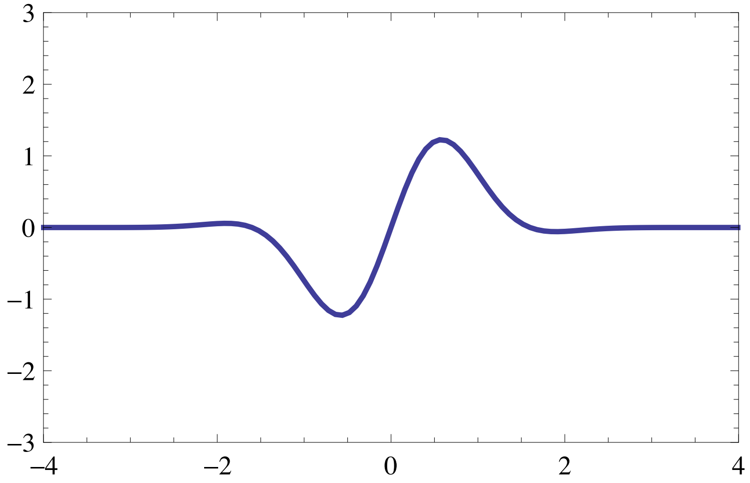
But harnessing the power of stochastic physical systems, that’s just cool!

Backup slides

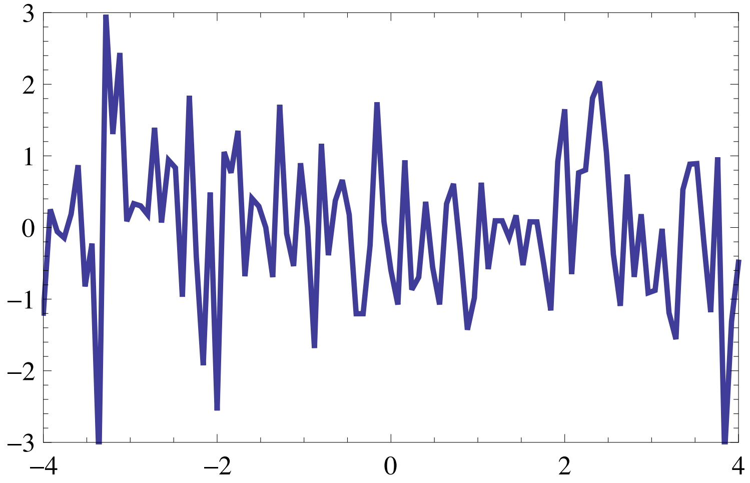
Michele Vallisneri, Jet Propulsion Laboratory

GW science in a nutshell:
GW detection with addition, subtraction, and multiplication

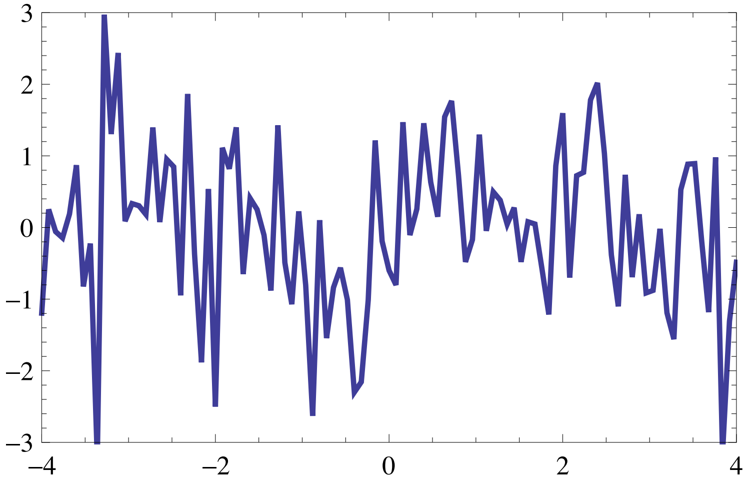
data = signal + noise



+

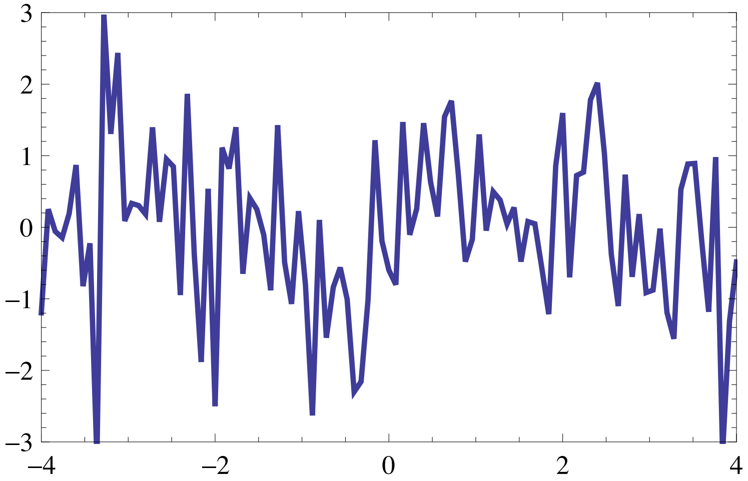


=



(SNR = 5)

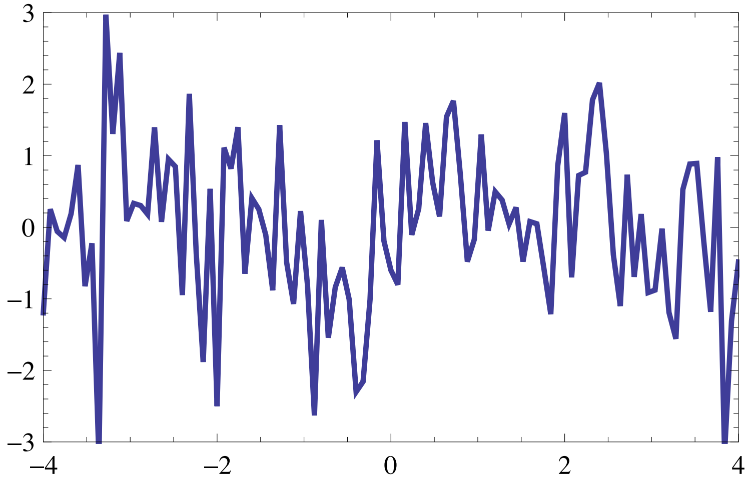
therefore: noise = data – signal; to assess detection, we ask which instance of noise is more probable?



(no signal)

or

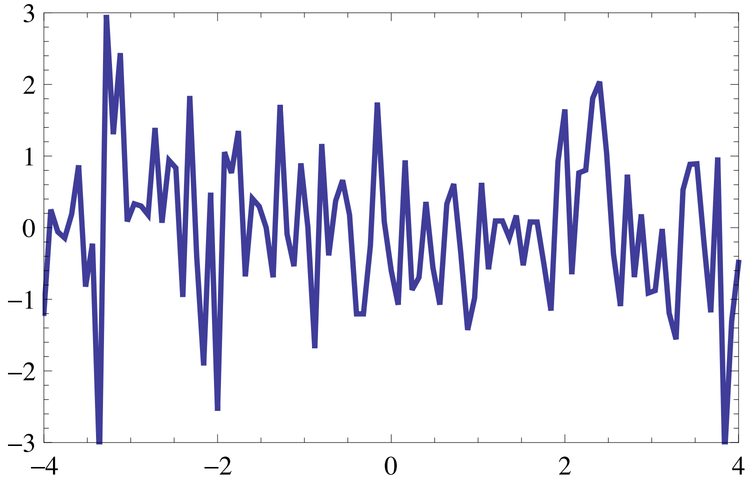
therefore: noise = data – signal; to assess detection, we ask which instance of noise is more probable?



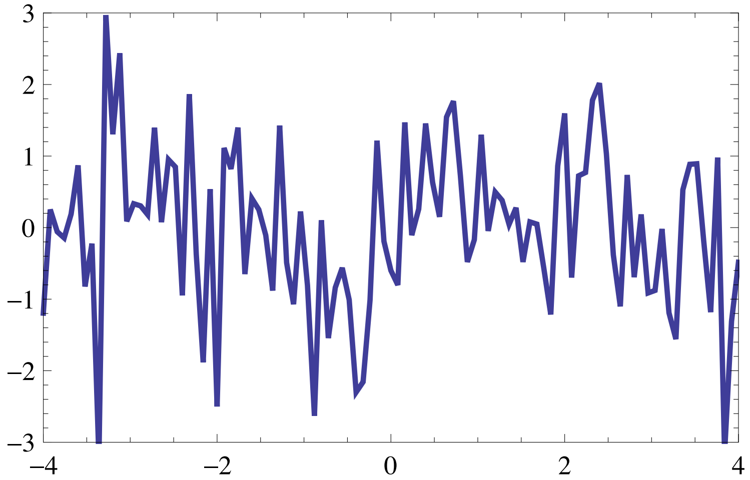
(no signal)

or

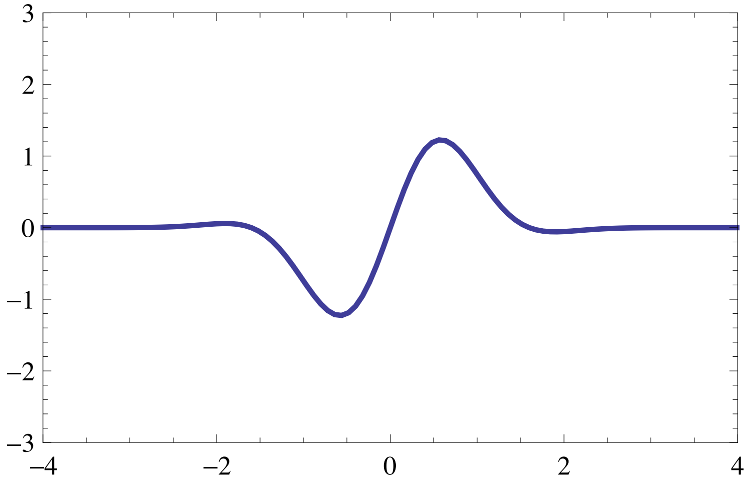
(signal hidden in noise, so we subtract it out)



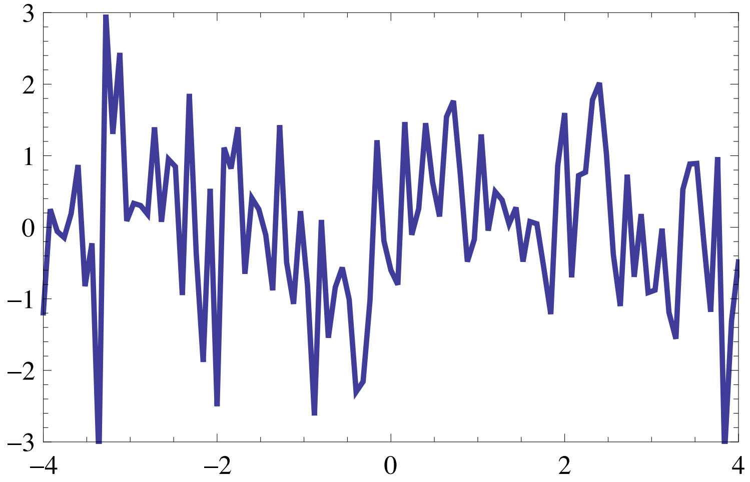
=



−



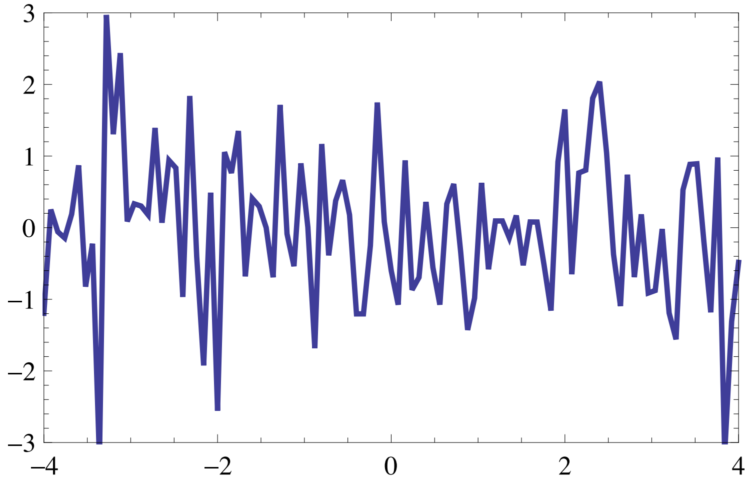
therefore: noise = data – signal; to assess detection, we ask which instance of noise is more probable?



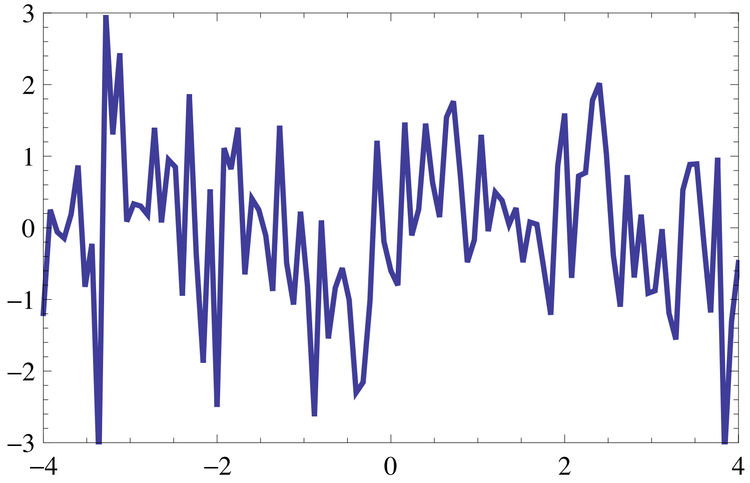
(no signal)

(signal hidden in noise, so we subtract it out)

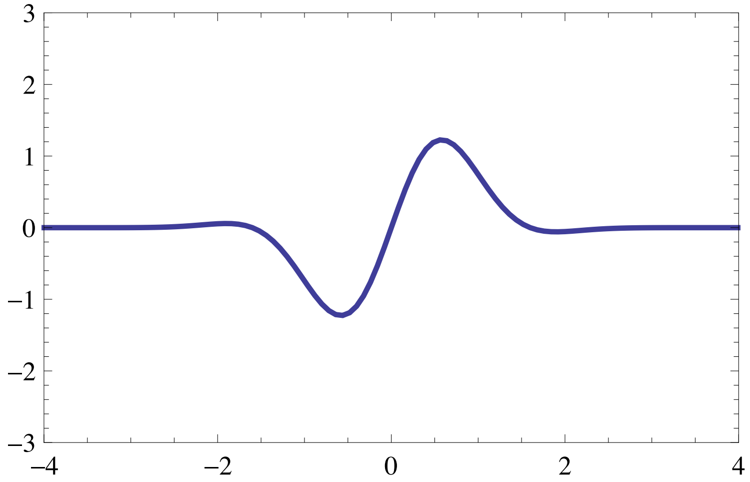
or



=

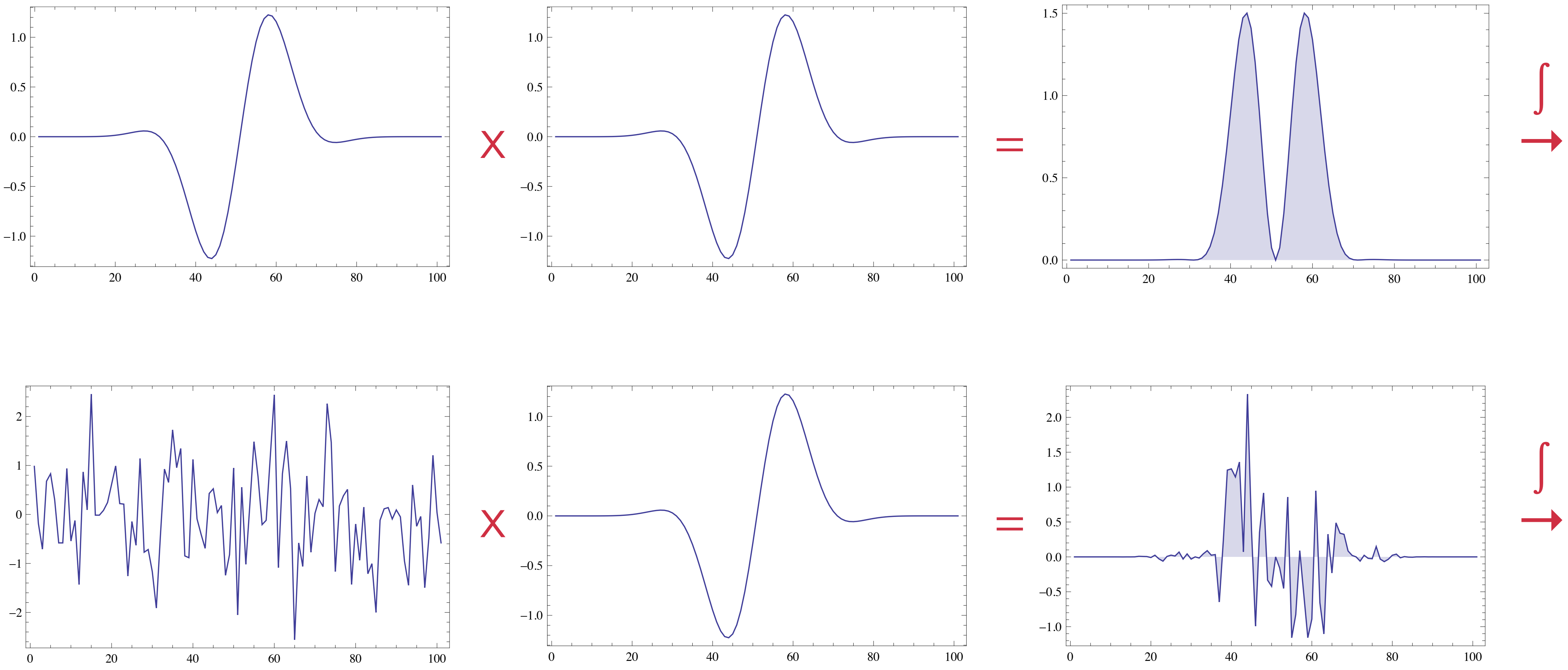


−



the ratio of probabilities is ~ exp SNR^2/2,
(here ~ 270,000)

an intuitive interpretation of this process is in terms of correlation products/matched filtering



∫ → 25

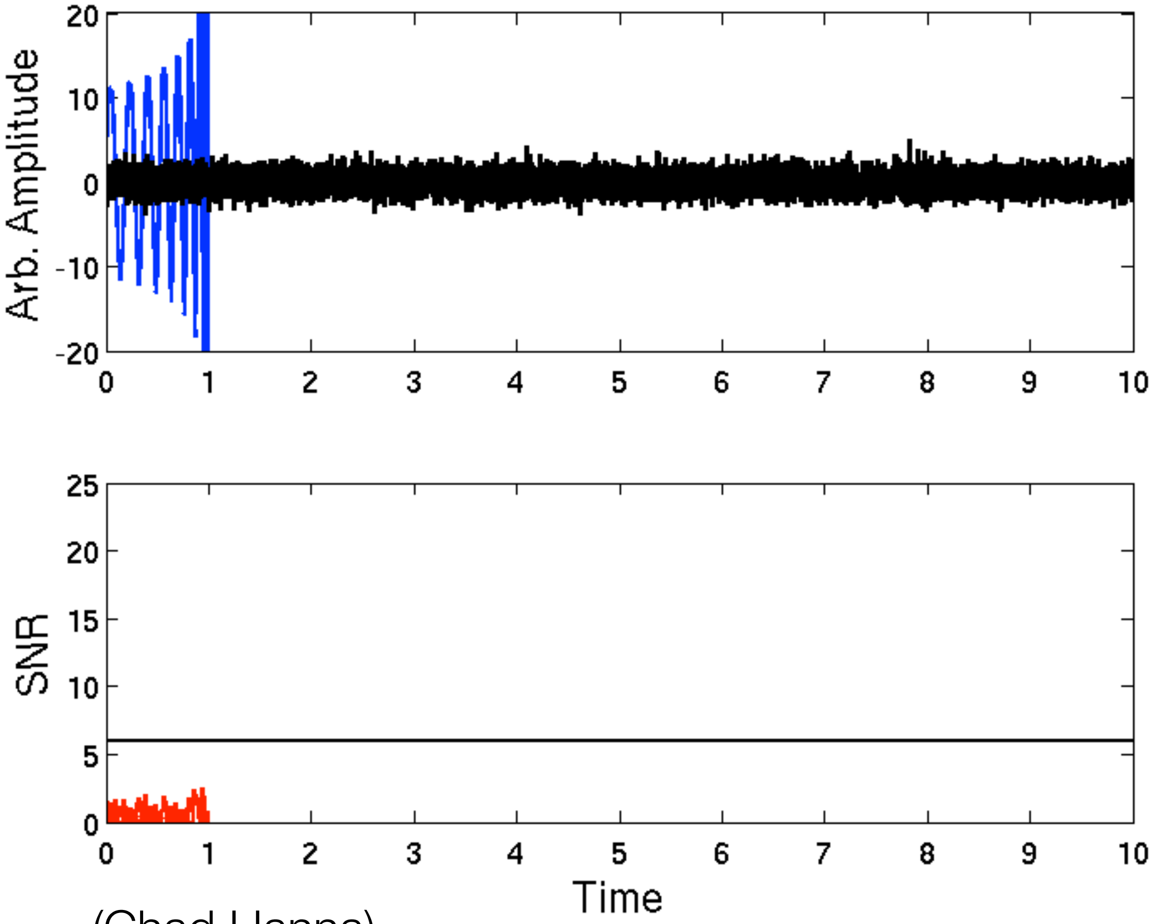
∫ → 3.39

an intuitive interpretation of this process
is in terms of correlation products/matched filtering

detector
data

$$\mathrm{SNR}(t_0) = \int_0^{\Delta t} s(t+t_0)h(t)dt$$

signal
“template”



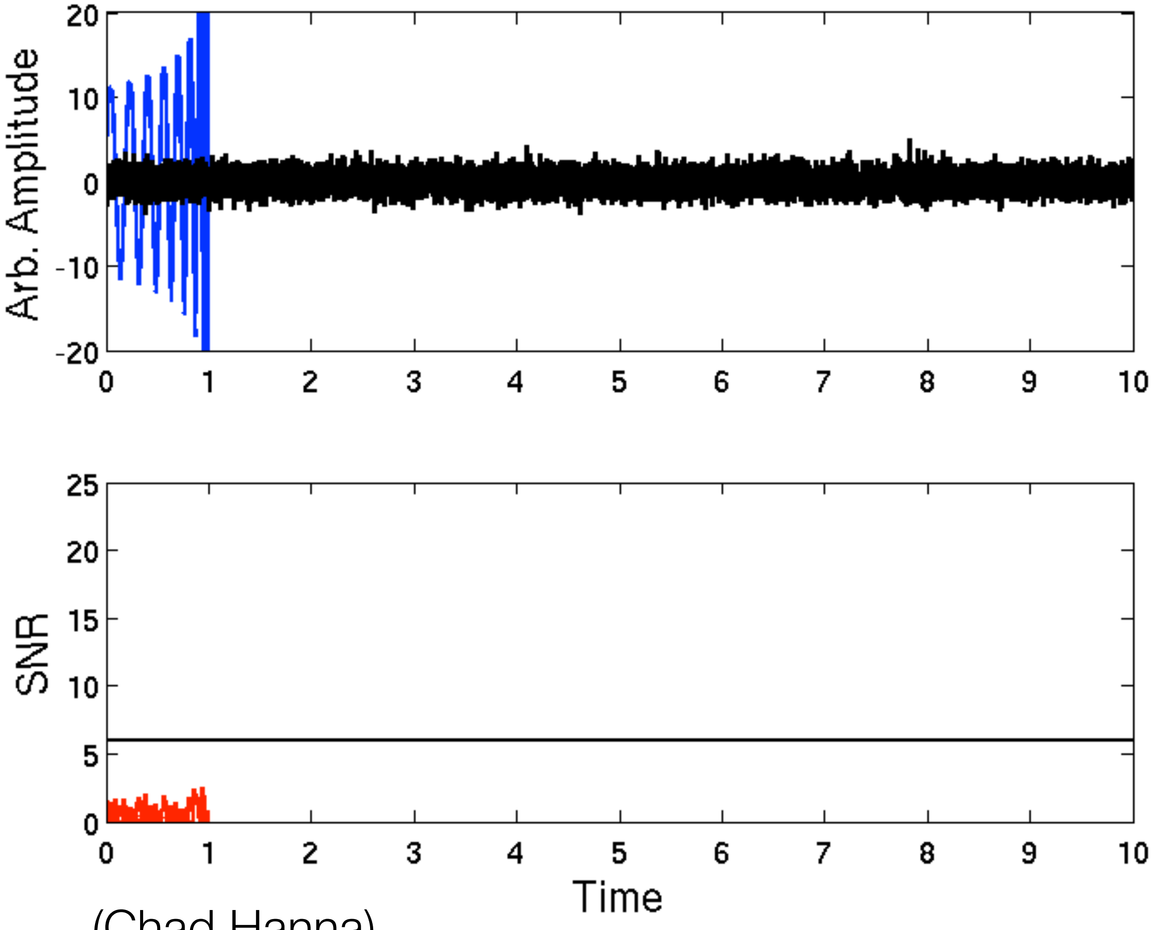
(Chad Hanna)

an intuitive interpretation of this process is in terms of correlation products/matched filtering

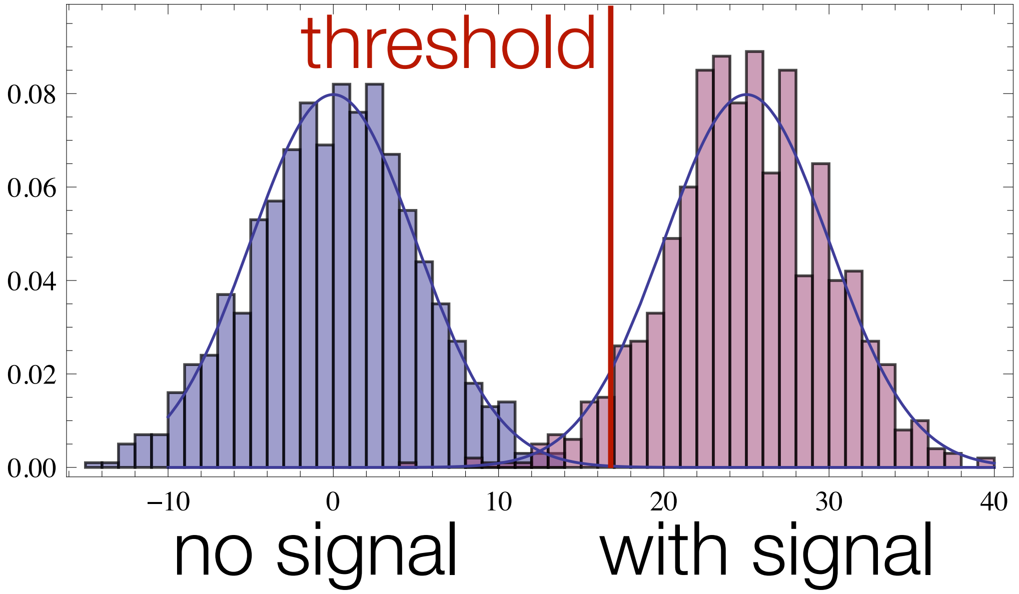
detector data

$$\mathrm{SNR}(t_0) = \int_0^{\Delta t} s(t + t_0) h(t) dt$$

signal "template"



(Chad Hanna)



GW detection in practice [see PRD 87, 024033, 2013]

filter detector output
with theoretical templates

GW detection in practice [see PRD 87, 024033, 2013]

condition and calibrate detector output

filter detector output with theoretical templates

GW detection in practice [see PRD 87, 024033, 2013]

condition and calibrate detector output

filter detector output with theoretical templates

request coincidence and consistency among detectors

GW detection in practice [see PRD 87, 024033, 2013]

condition and calibrate detector output

filter detector output with theoretical templates

request coincidence and consistency among detectors

apply data-quality cuts and signal vetos

GW detection in practice [see PRD 87, 024033, 2013]

condition and calibrate detector output

filter detector output with theoretical templates

request coincidence and consistency among detectors

apply data-quality cuts and signal vetos

estimate statistical significance

(estimate background, using coincidence between time slides)

GW detection in practice [see PRD 87, 024033, 2013]

condition and calibrate detector output

filter detector output with theoretical templates

request coincidence and consistency among detectors

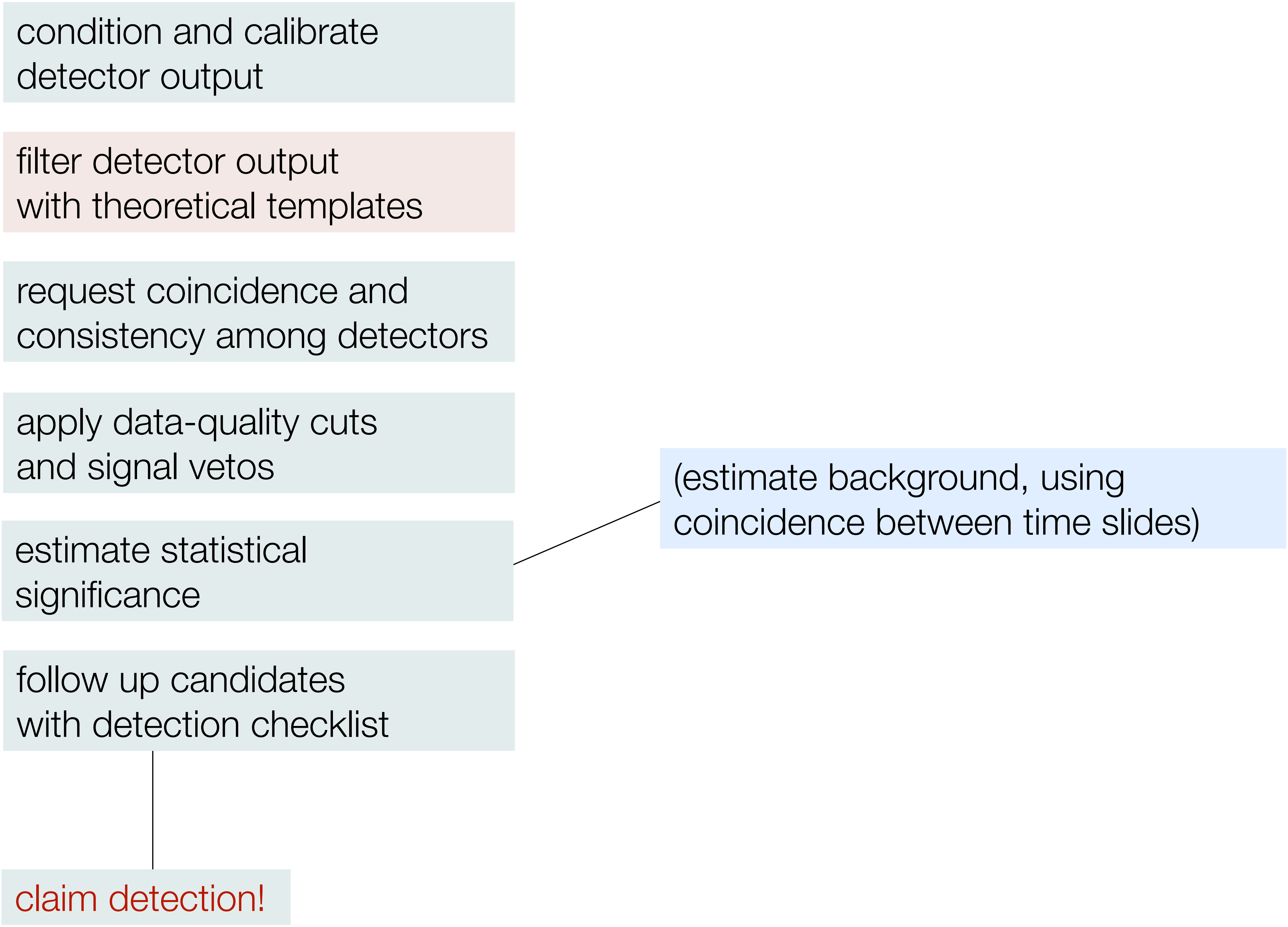
apply data-quality cuts and signal vetos

estimate statistical significance

(estimate background, using coincidence between time slides)

follow up candidates with detection checklist

GW detection in practice [see PRD 87, 024033, 2013]



condition and calibrate detector output

filter detector output with theoretical templates

request coincidence and consistency among detectors

apply data-quality cuts and signal vetos

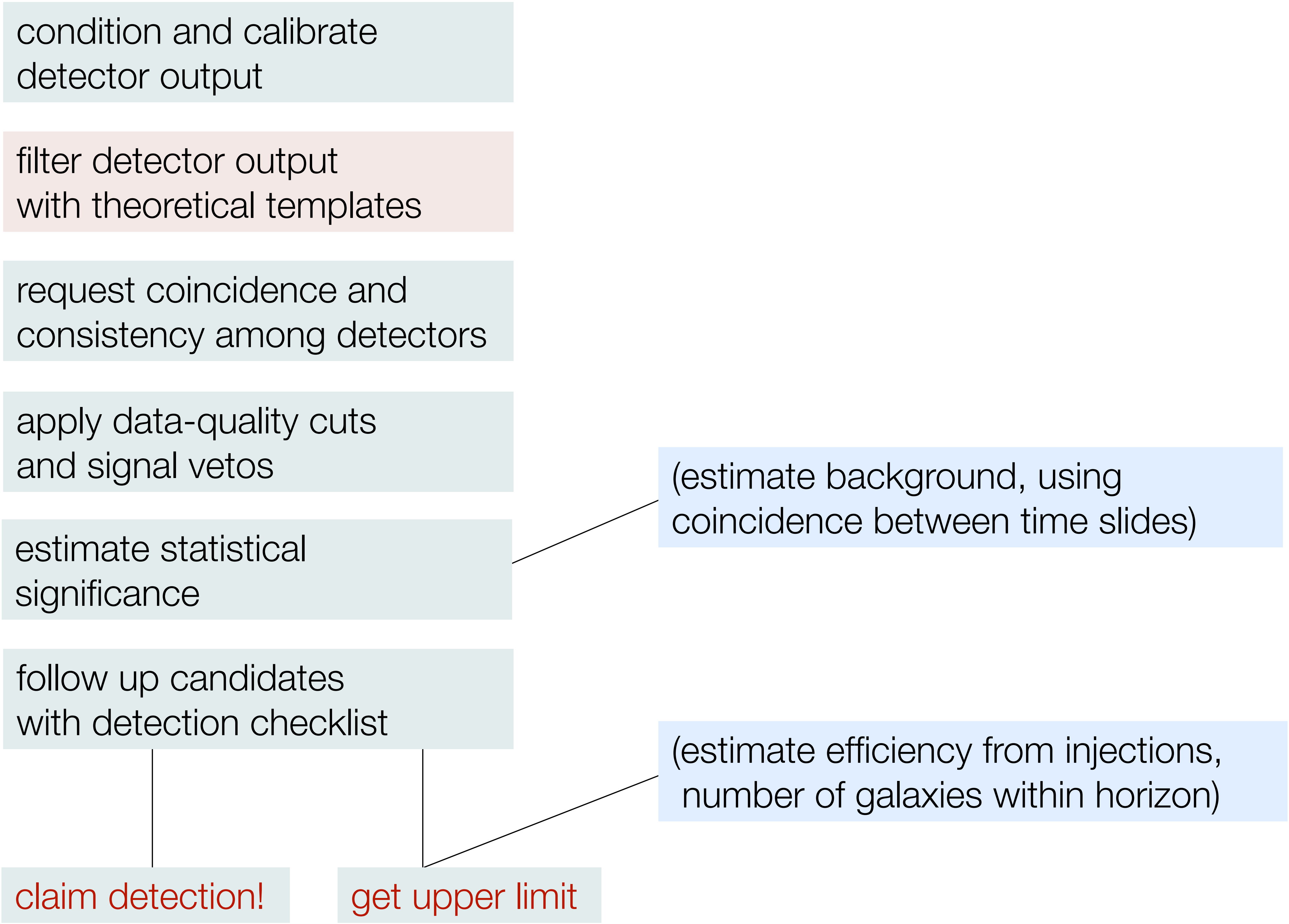
estimate statistical significance

(estimate background, using coincidence between time slides)

follow up candidates with detection checklist

claim detection!

GW detection in practice [see PRD 87, 024033, 2013]



condition and calibrate detector output

filter detector output with theoretical templates

request coincidence and consistency among detectors

apply data-quality cuts and signal vetos

estimate statistical significance

(estimate background, using coincidence between time slides)

follow up candidates with detection checklist

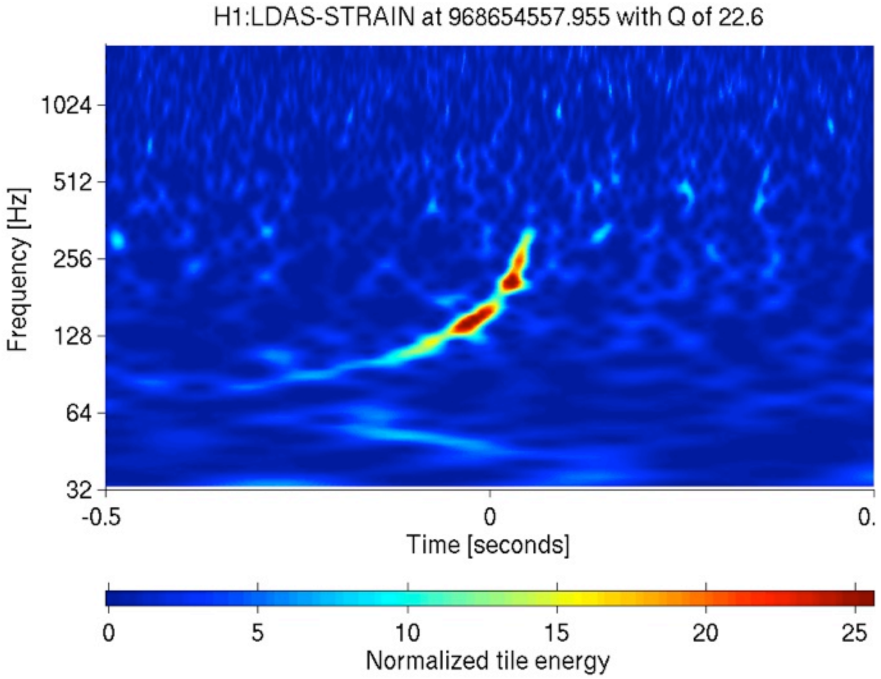
(estimate efficiency from injections, number of galaxies within horizon)

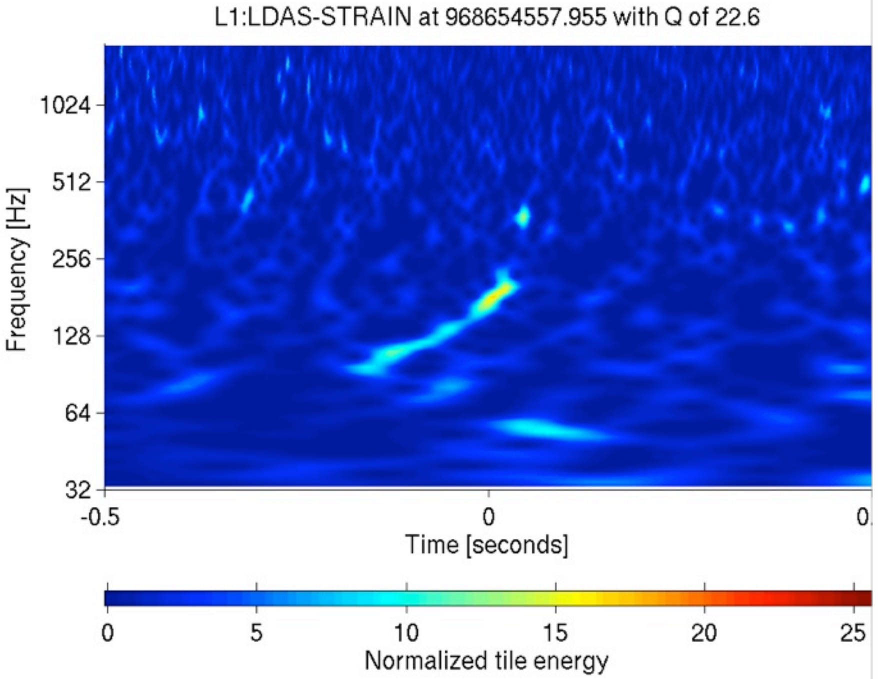
claim detection!

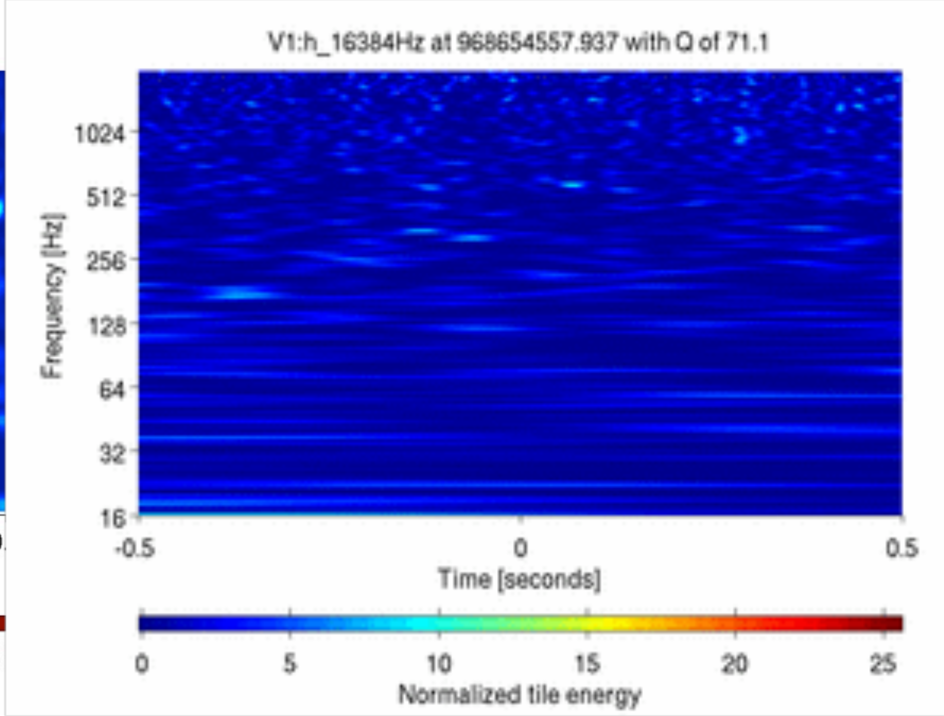
get upper limit

GW100916, the Big Dog

GW100916, the Big Dog

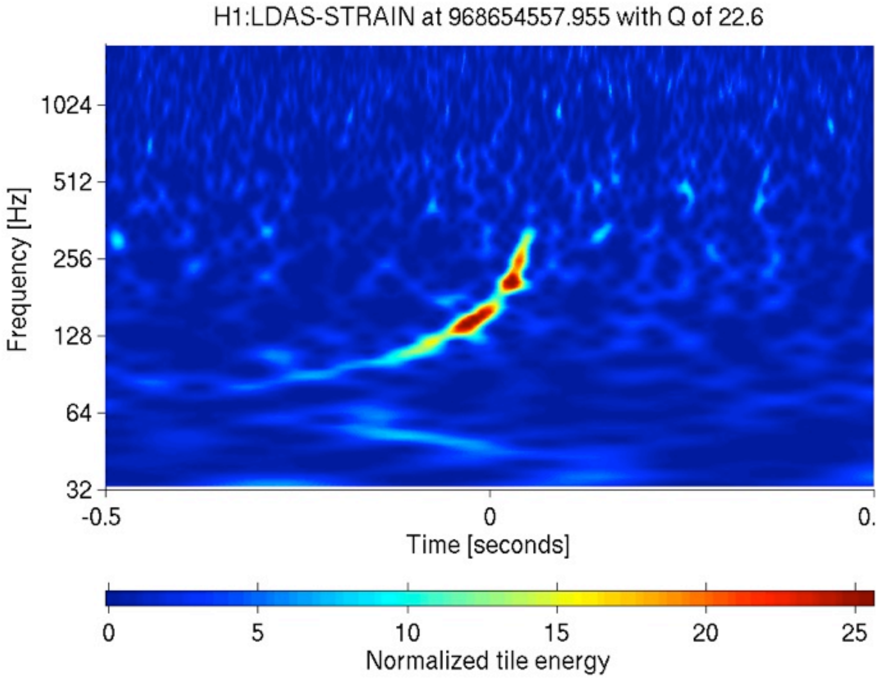

H1:LDAS-STRAIN at 968654557.955 with Q of 22.6

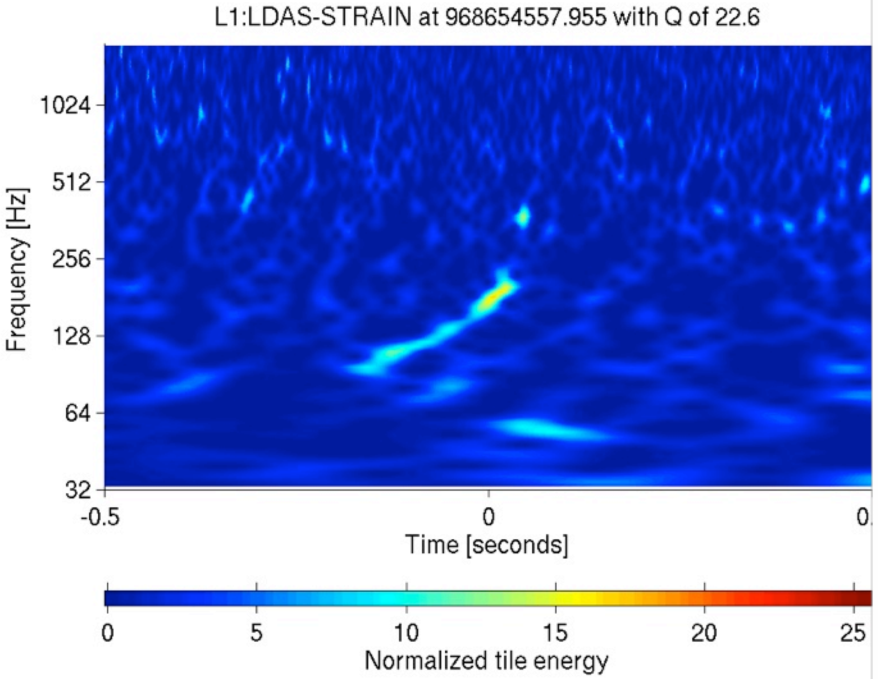

L1:LDAS-STRAIN at 968654557.955 with Q of 22.6

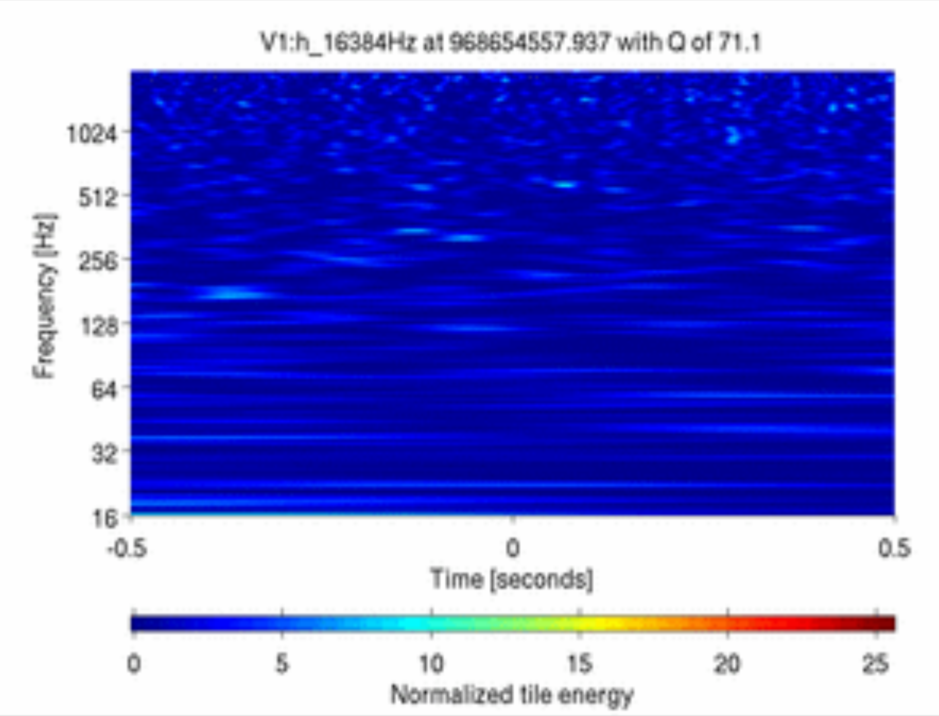

V1:h_16384Hz at 968654557.937 with Q of 71.1

GW100916, the Big Dog

- the LVC toasted with champagne before opening the envelope

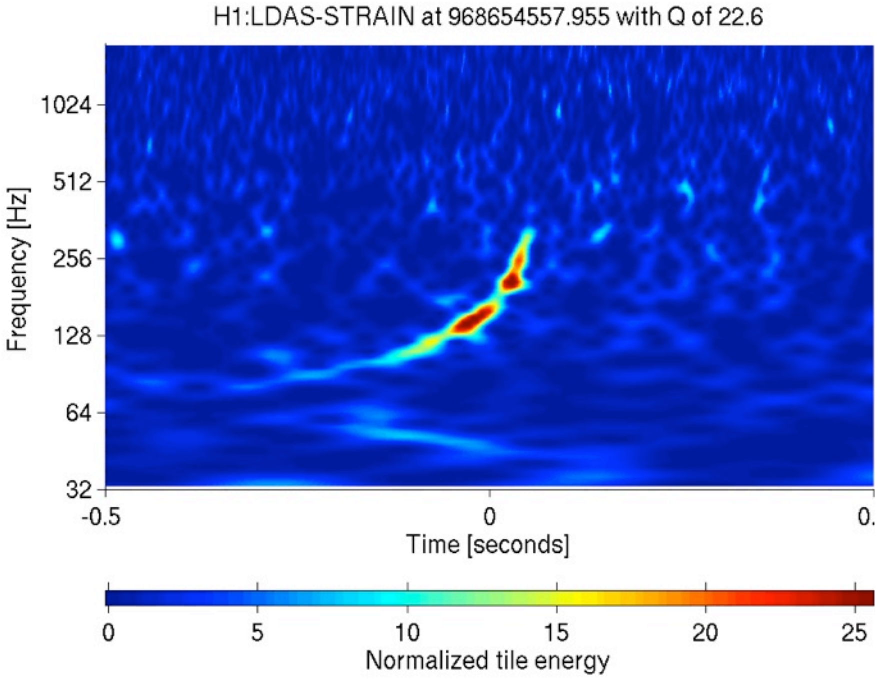


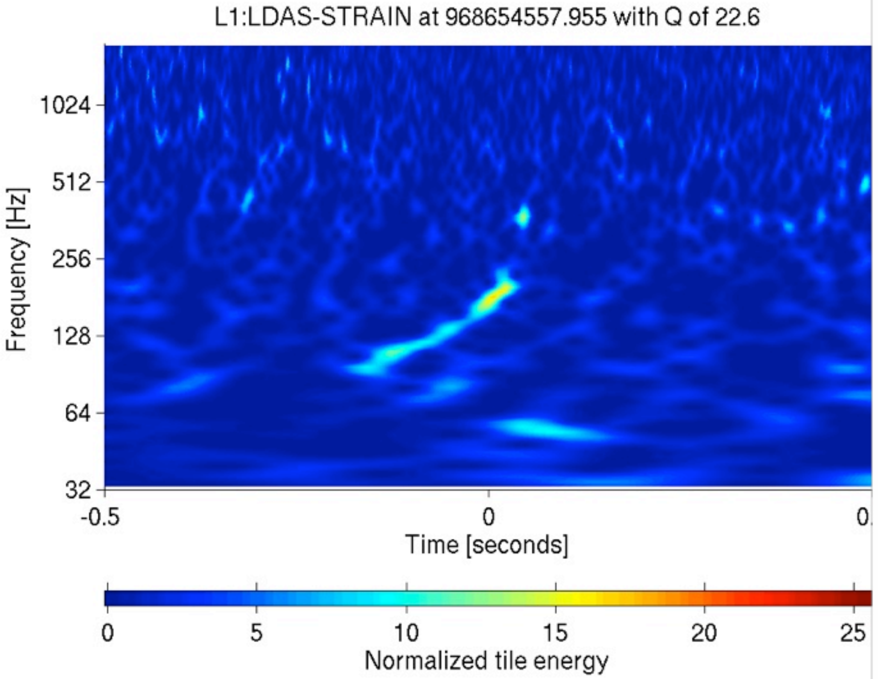


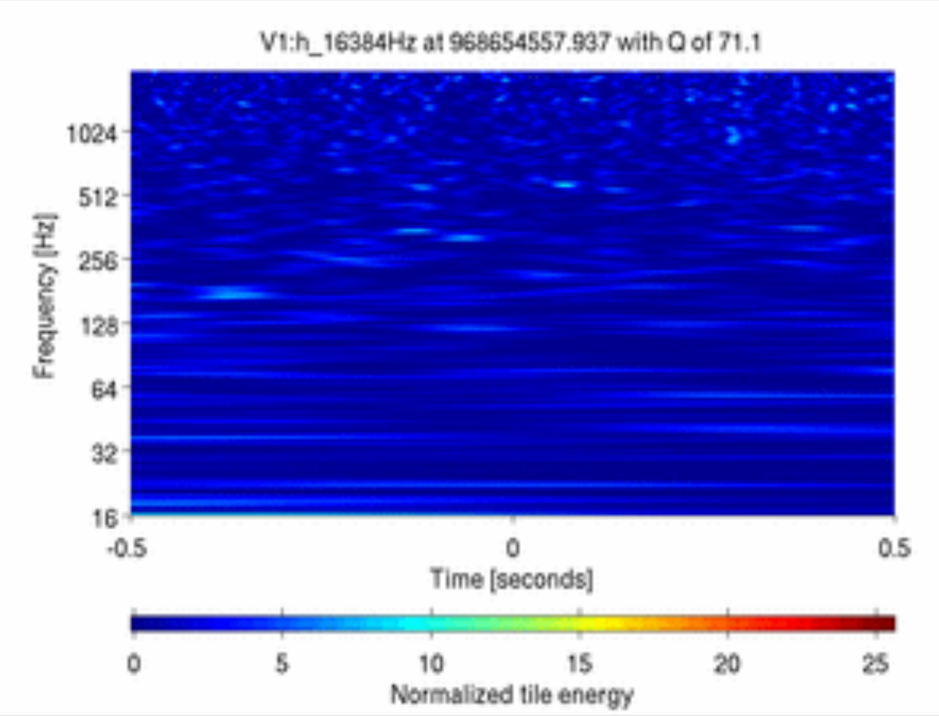


GW100916, the Big Dog

- the LVC toasted with champagne before opening the envelope
- unfortunately, it was a blind injection...

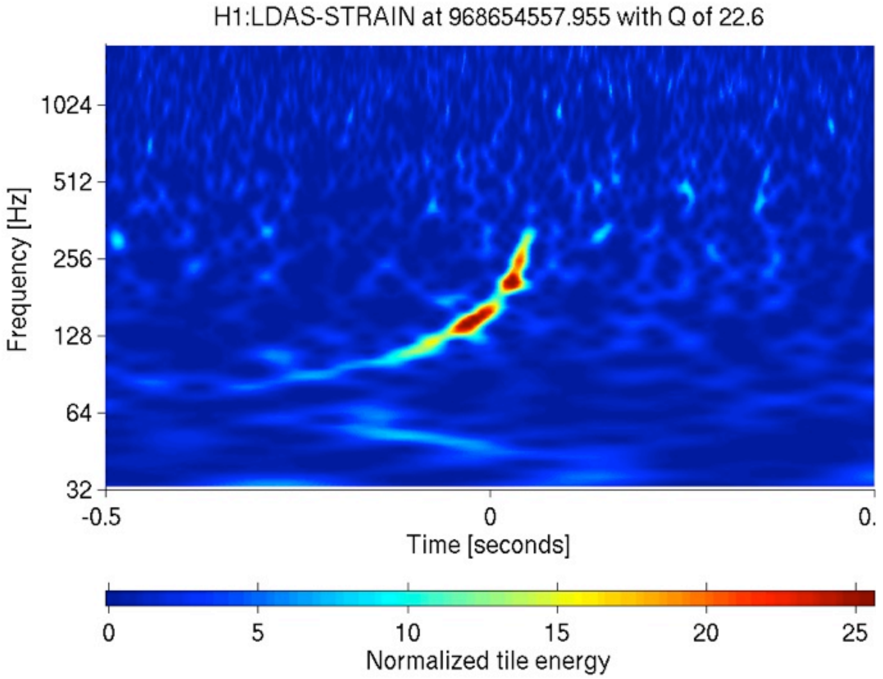

H1:LDAS-STRAIN at 968654557.955 with Q of 22.6

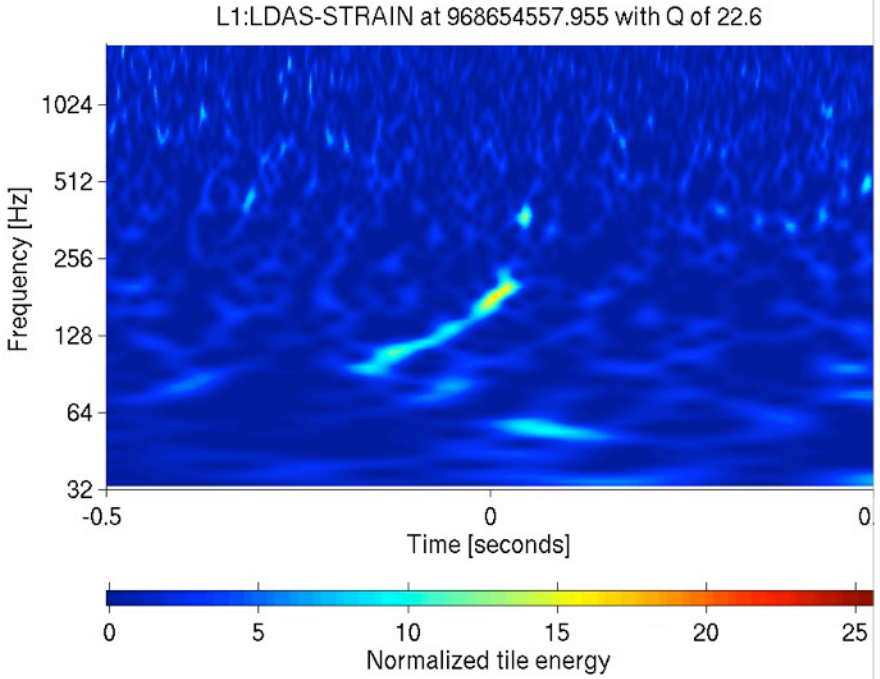

L1:LDAS-STRAIN at 968654557.955 with Q of 22.6

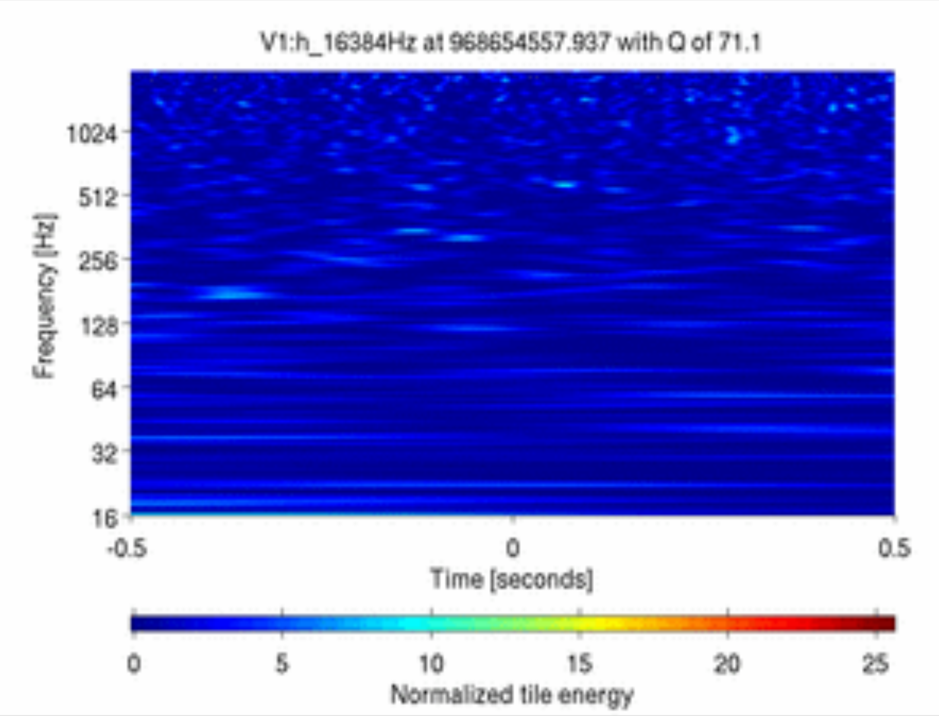

V1:h_16384Hz at 968654557.937 with Q of 71.1

GW100916, the Big Dog

- the LVC toasted with champagne before opening the envelope
- unfortunately, it was a blind injection...
- ...but we found it!

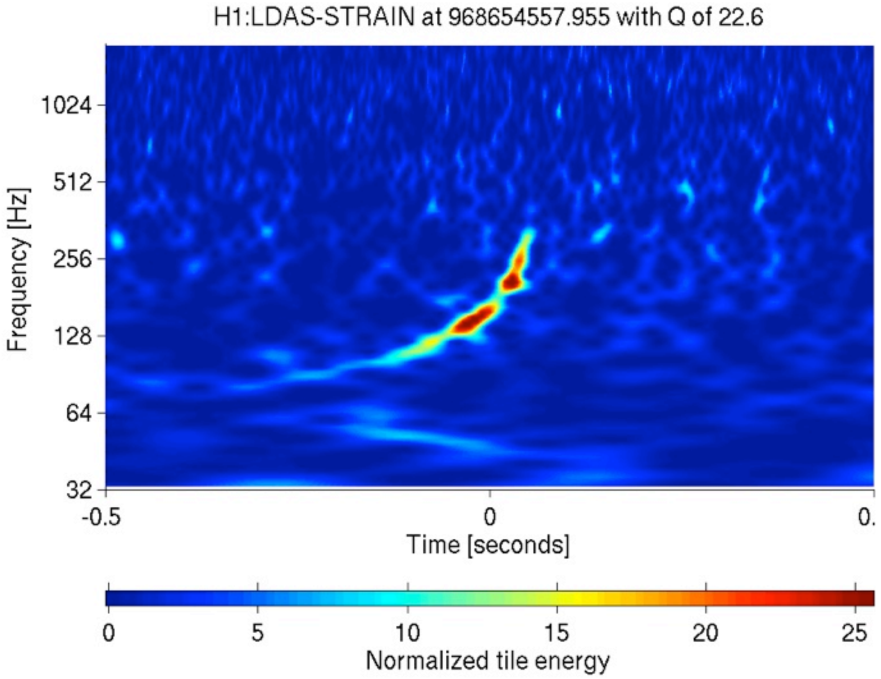


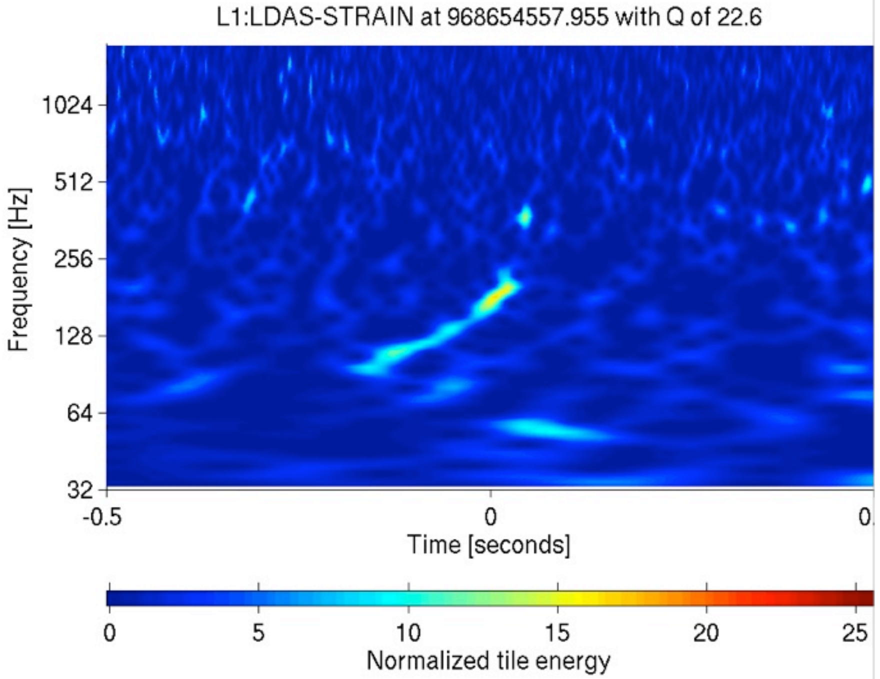


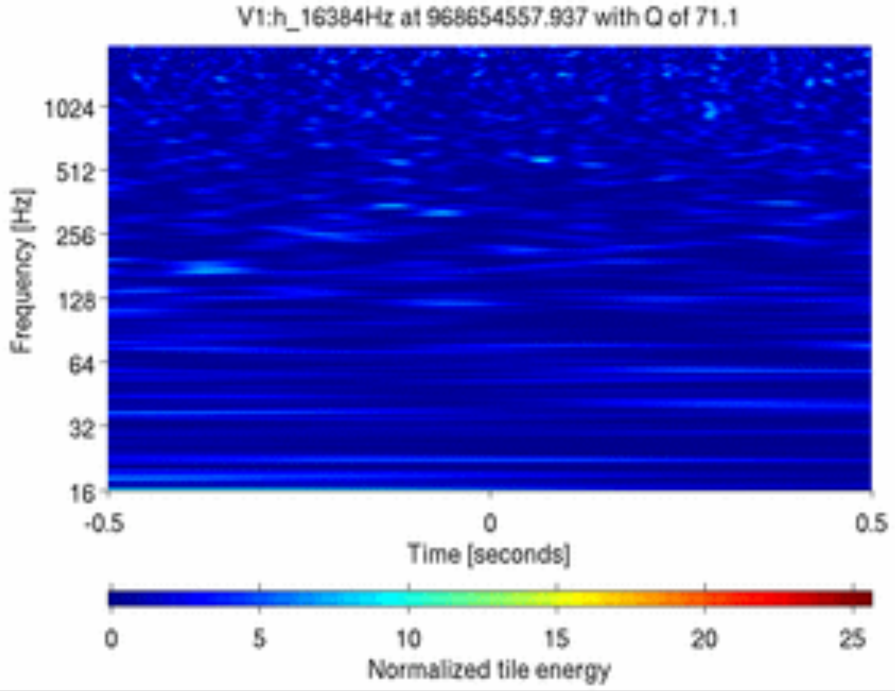


GW100916, the Big Dog

- the LVC toasted with champagne before opening the envelope
- unfortunately, it was a blind injection...
- ...but we found it!
- the process exercised methods, protocols, and people

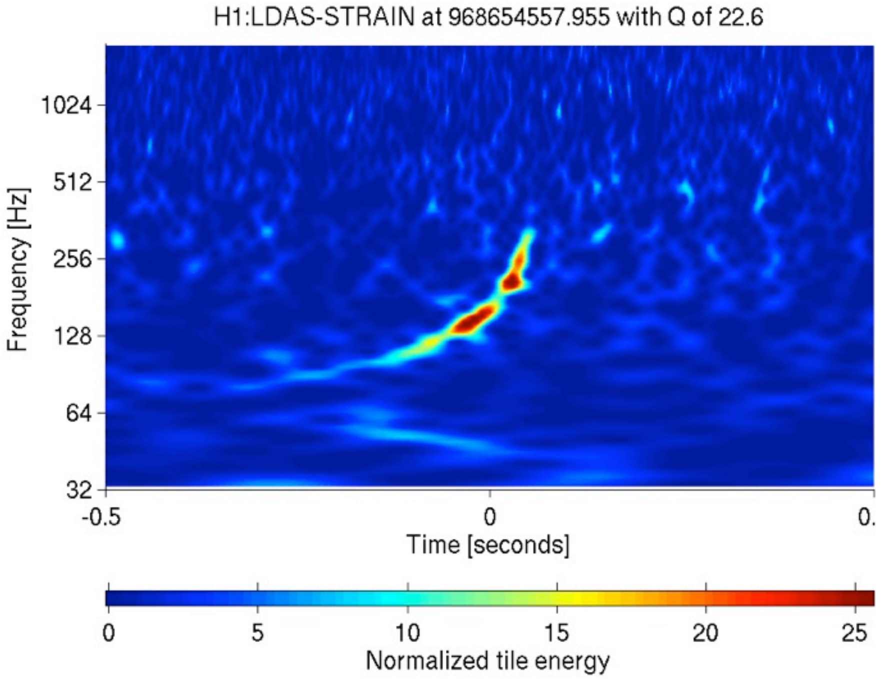


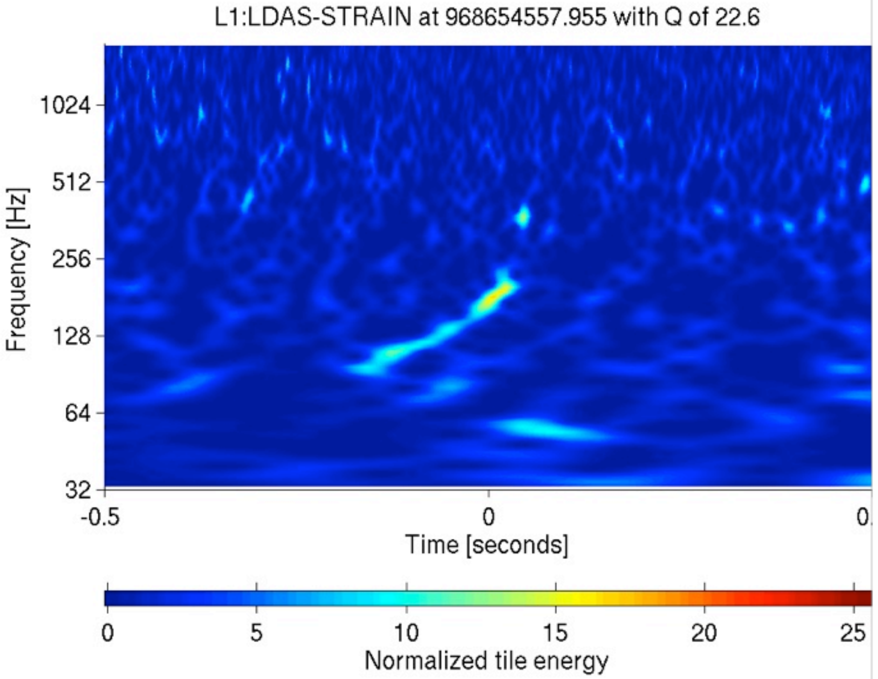


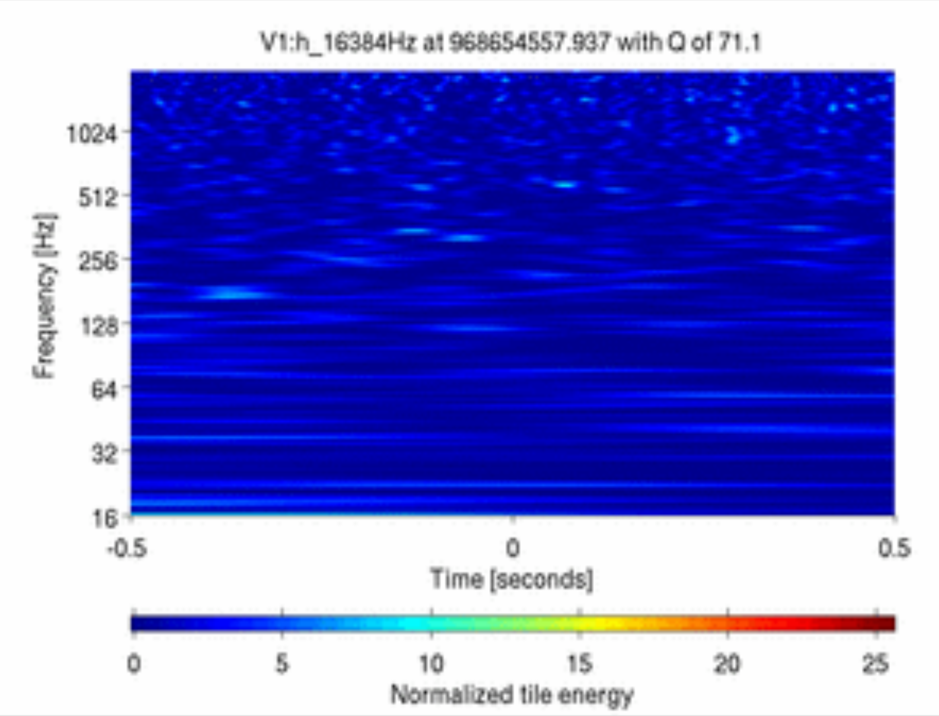


GW100916, the Big Dog

- the LVC toasted with champagne before opening the envelope
- unfortunately, it was a blind injection...
- ...but we found it!
- the process exercised methods, protocols, and people
- it showed the perils of theory, experiment, software







we design a decision scheme ("AG or GR?") with the Bayesian odds ratio $\mathcal{O}$ as the detection statistic; we set a threshold $\mathcal{O}^*$ and claim detection when $\mathcal{O} > \mathcal{O}^*$

model priors

parameter priors

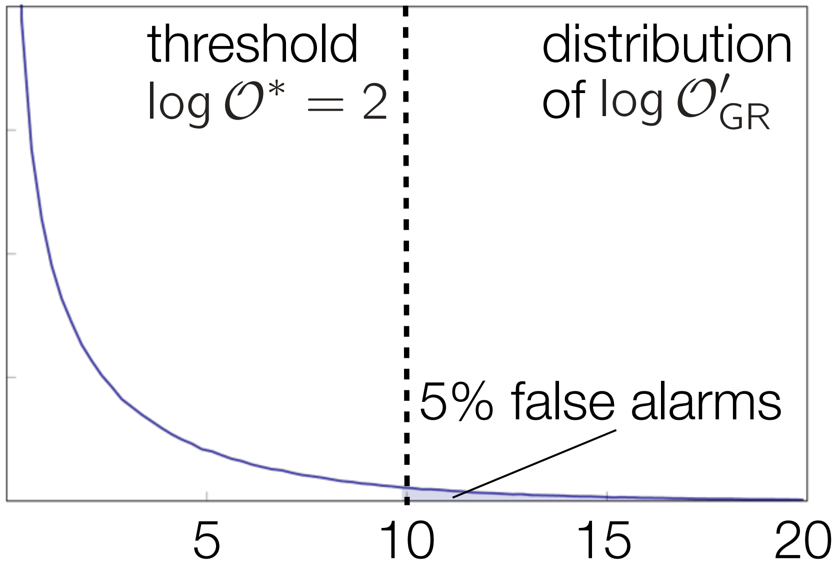
$$\mathcal{O} = \frac{P(\mathrm{AG}|s)}{P(\mathrm{GR}|s)} = \frac{P(\mathrm{AG}) \int p(s|\theta^{i,a})\, p(\theta^{i,a})\, d\theta^{i,a}}{P(\mathrm{GR}) \int p(s|\theta^{i})\, p(\theta^{i})\, d\theta^{i}}$$

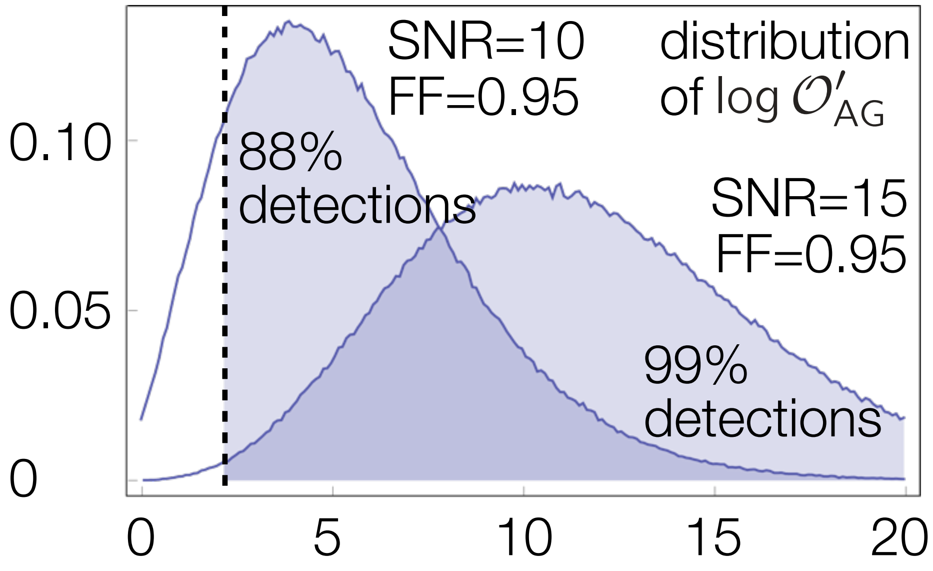
AG parameters

GR parameters

evidence (= marginal likelihood)
for AG and GR models

likelihood





background: true signal is GR

detection efficiency: true signal is AG

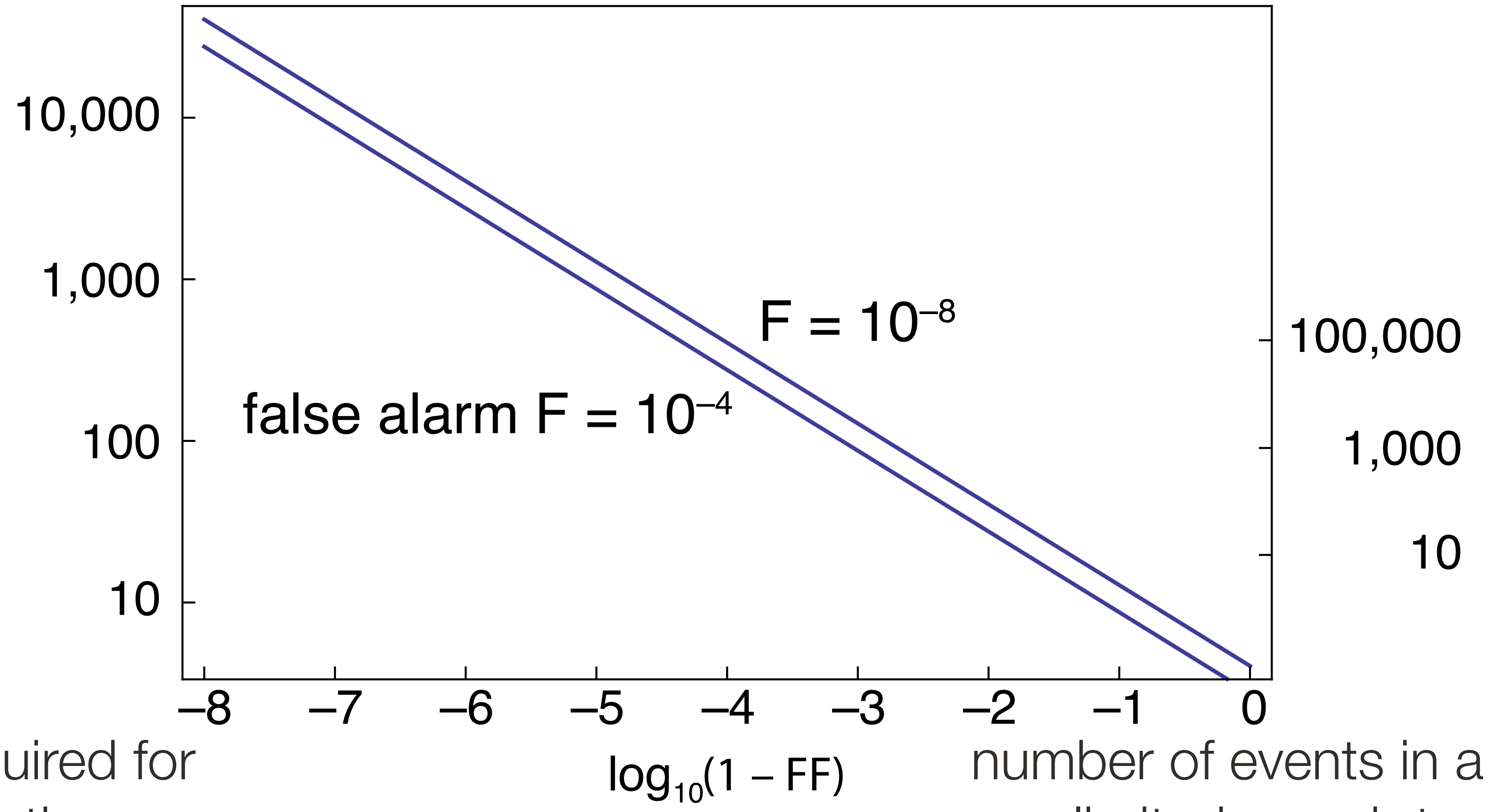
$$\mathcal{O}'_{\mathrm{GR}} = e^{x^2/2}$$

$$\mathcal{O}'_{\mathrm{AG}} = e^{x^2/2 + x\sqrt{2(1-\mathrm{FF})}\,\mathrm{SNR} + (1-\mathrm{FF})\,\mathrm{SNR}^2}$$

renormalized odds ratios (model priors
and Occam factors cancel out, see paper)

x is a normal random variable with zero mean
and unit variance (a function on noise realization)

for strong signals, $\mathcal{O}'_{\rm GR}$ and $\mathcal{O}'_{\rm AG}$ are remarkably simple functions of FF and SNR alone. For a fixed false-alarm rate, we then ask what SNR yields 50%-efficient AG detection, as a function of FF



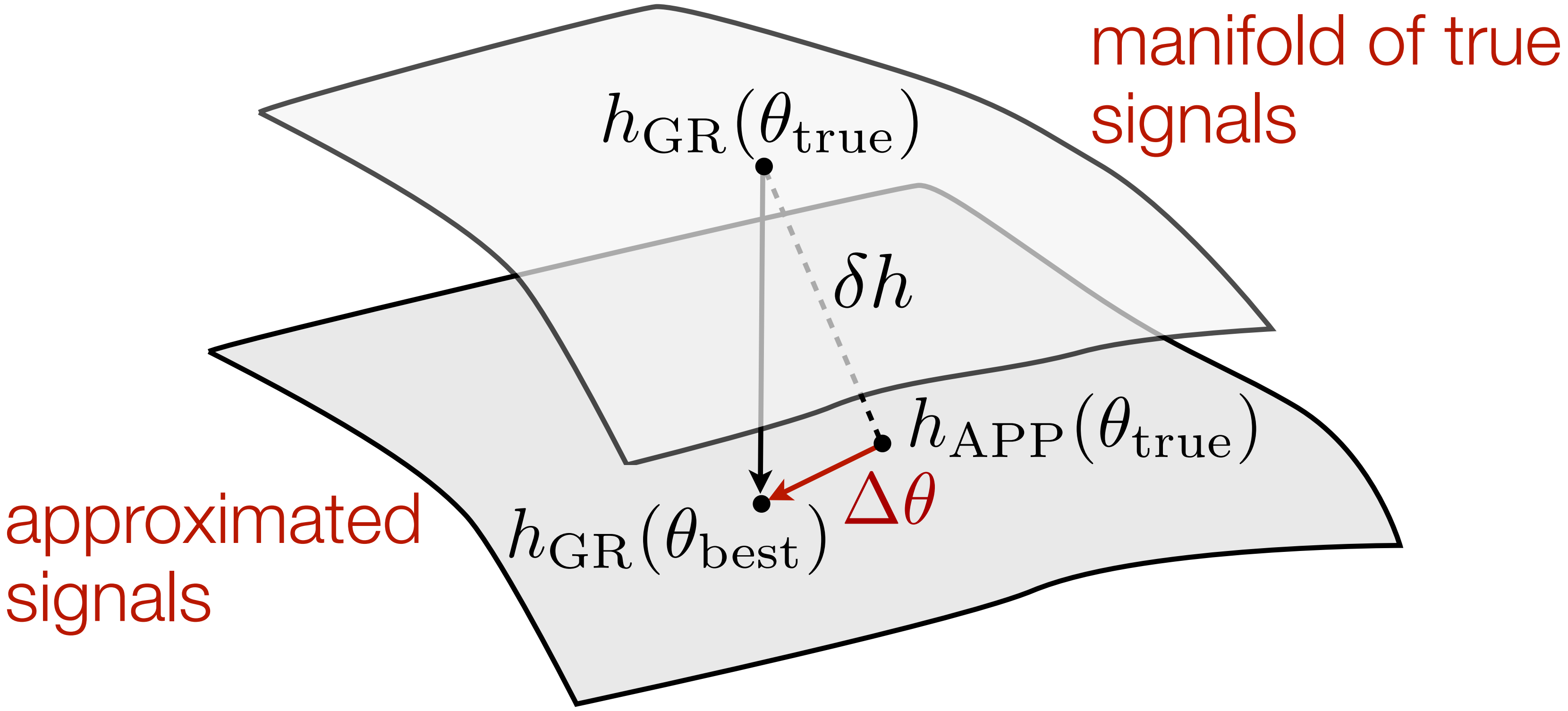
SNR required for AG detection with 50% efficiency

number of events in a volume-limited search to yield that SNR as the loudest event

only very strong AG effects (FF of 0.9–0.99) would be seen in volume-limited searches, so GR tests may have to wait for third-generation ground-based detectors, or for space detectors

an application [Vallisneri and Yunes, arXiv/1301.2627]: “fundamental” bias versus the detection of modified gravity

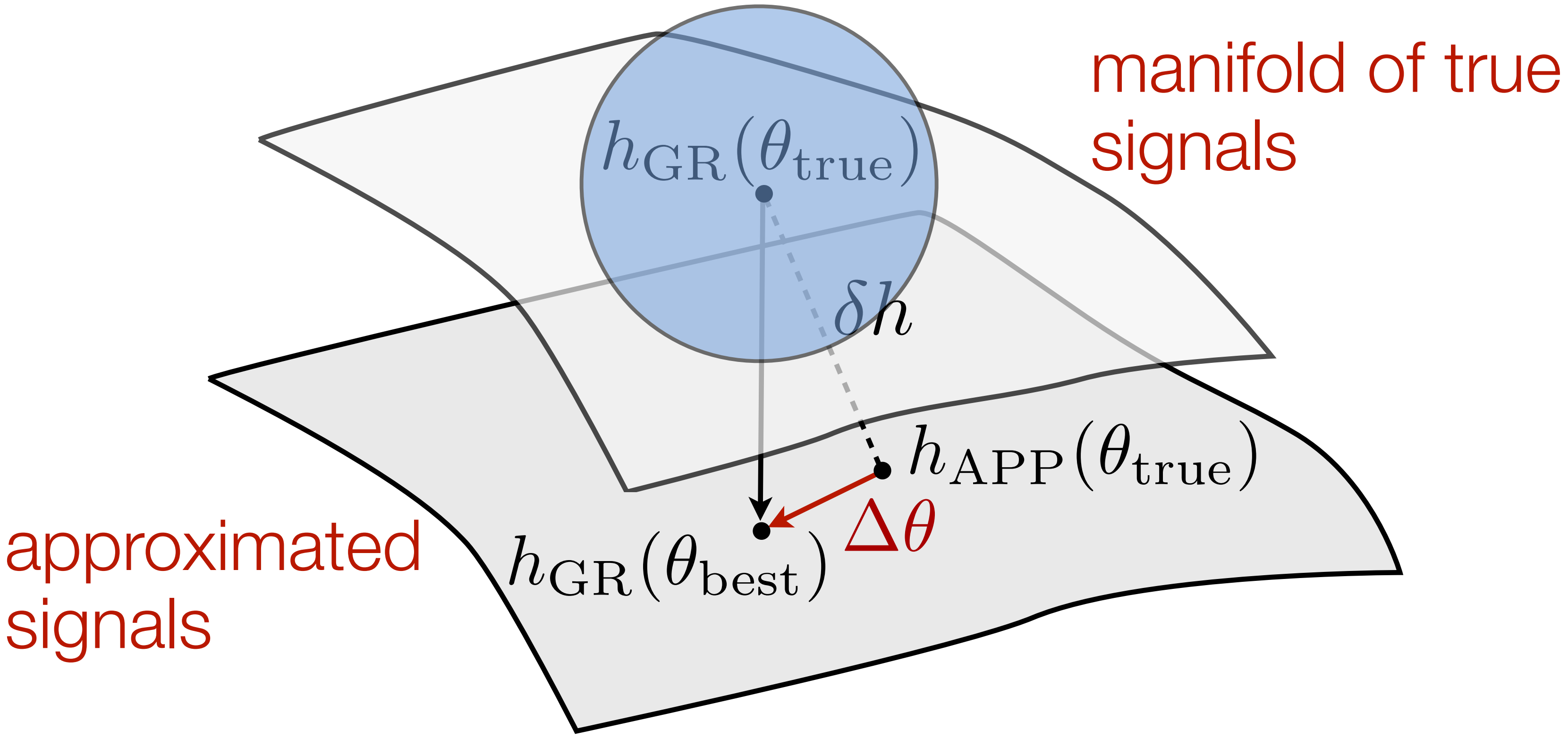
if model signals (GR) differ from true signals (MG) with the same parameters, the best-fitting template will be displaced by theoretical error, which is SNR-independent [see Cutler & Vallisneri 2007]


manifold of true signals
$h_{\mathrm{GR}}(\theta_{\mathrm{true}})$
δh
$h_{\mathrm{APP}}(\theta_{\mathrm{true}})$
$\Delta\theta$
approximated signals
$h_{\mathrm{GR}}(\theta_{\mathrm{best}})$

$$\delta\vec{\theta}_{\mathrm{th}} = (F_{\mathrm{bf}}^{-1})^{\alpha\beta}\big(h_{,\beta}(\vec{\theta}_{\mathrm{bf}})\big|\delta h(\vec{\theta}_{\mathrm{bf}})\big)$$

an application [Vallisneri and Yunes, arXiv/1301.2627]: "fundamental" bias versus the detection of modified gravity

if model signals (GR) differ from true signals (MG) with the same parameters, the best-fitting template will be displaced by theoretical error, which is SNR-independent [see Cutler & Vallisneri 2007]


manifold of true signals
$h_{\mathrm{GR}}(\theta_{\mathrm{true}})$
δh
$h_{\mathrm{APP}}(\theta_{\mathrm{true}})$
$\Delta\theta$
approximated signals
$h_{\mathrm{GR}}(\theta_{\mathrm{best}})$

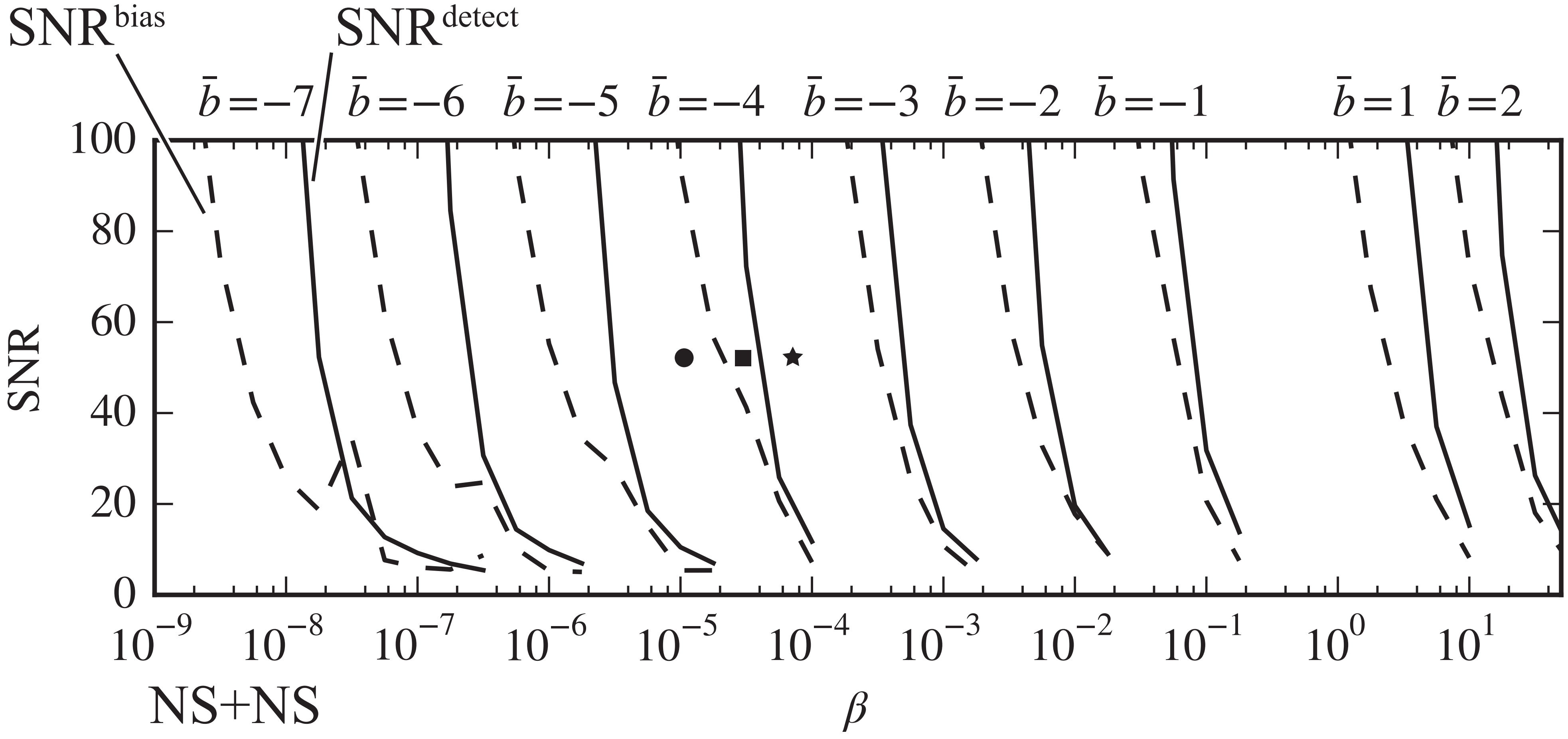
$$\delta\vec{\theta}_{\mathrm{th}} = (F_{\mathrm{bf}}^{-1})^{\alpha\beta}\big(h_{,\beta}(\vec{\theta}_{\mathrm{bf}})\big|\delta h(\vec{\theta}_{\mathrm{bf}})\big)$$

representing modified gravity as “parametrized post-Einstein,” we compare the MG SNR$^{\text{detect}}$ with the SNR$^{\text{bias}}$ where δθ$_{\text{th}}$ > δθ$_{\text{stat}}$

$$h_{\mathrm{GR}}(f) = A_{\mathrm{GR}}(f) e^{i\Psi_{\mathrm{GR}}(f)} \quad \delta\Psi(f) = \beta\, u^b$$

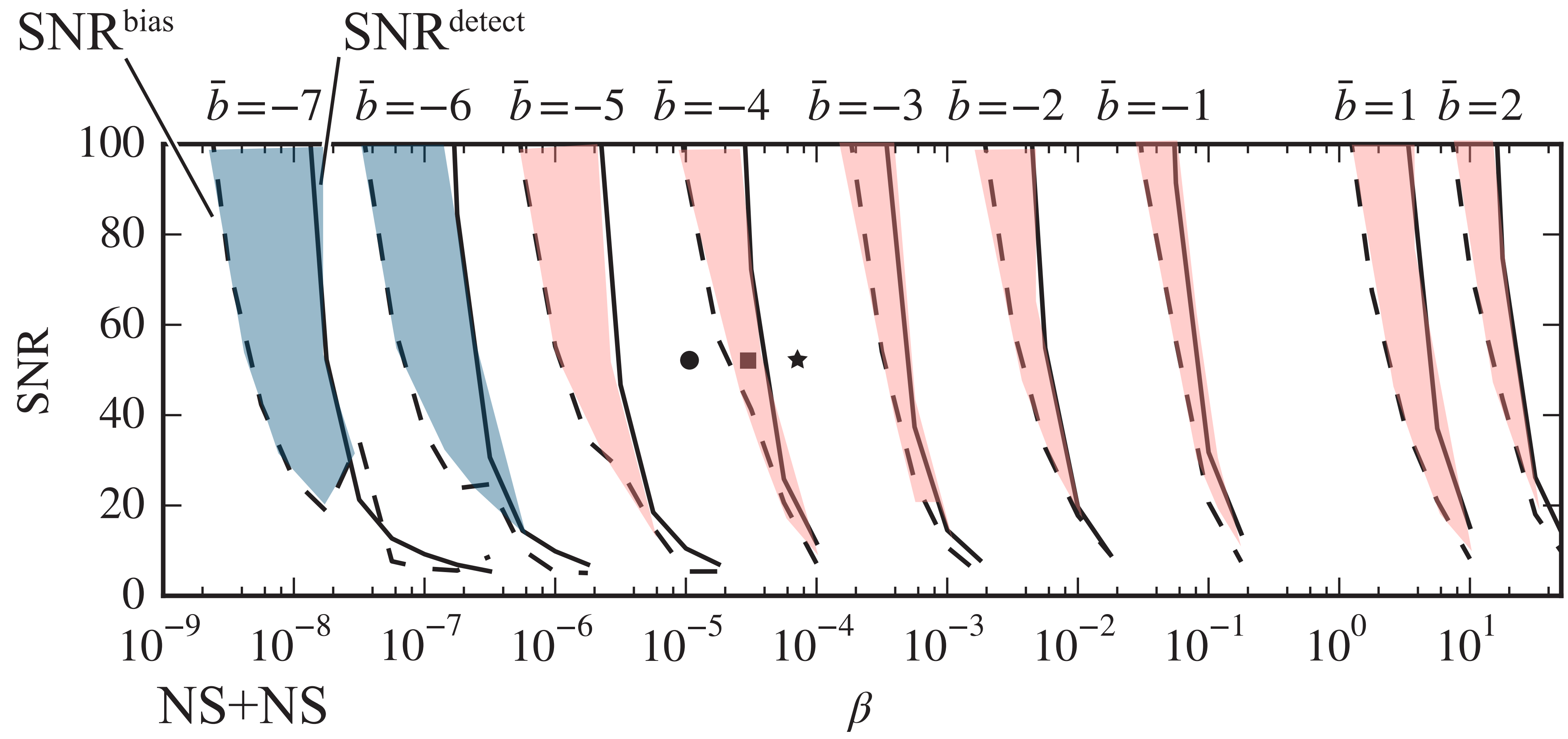
representing modified gravity as “parametrized post-Einstein,” we compare the MG SNR$^{\text{detect}}$ with the SNR$^{\text{bias}}$ where $\delta\theta_{\text{th}} > \delta\theta_{\text{stat}}$

$$h_{\rm GR}(f) = A_{\rm GR}(f) e^{i\Psi_{\rm GR}(f)} \quad \delta\Psi(f) = \beta\, u^b$$



representing modified gravity as "parametrized post-Einstein," we compare the MG SNR^detect^ with the SNR^bias^ where $\delta\theta_{\mathrm{th}} > \delta\theta_{\mathrm{stat}}$

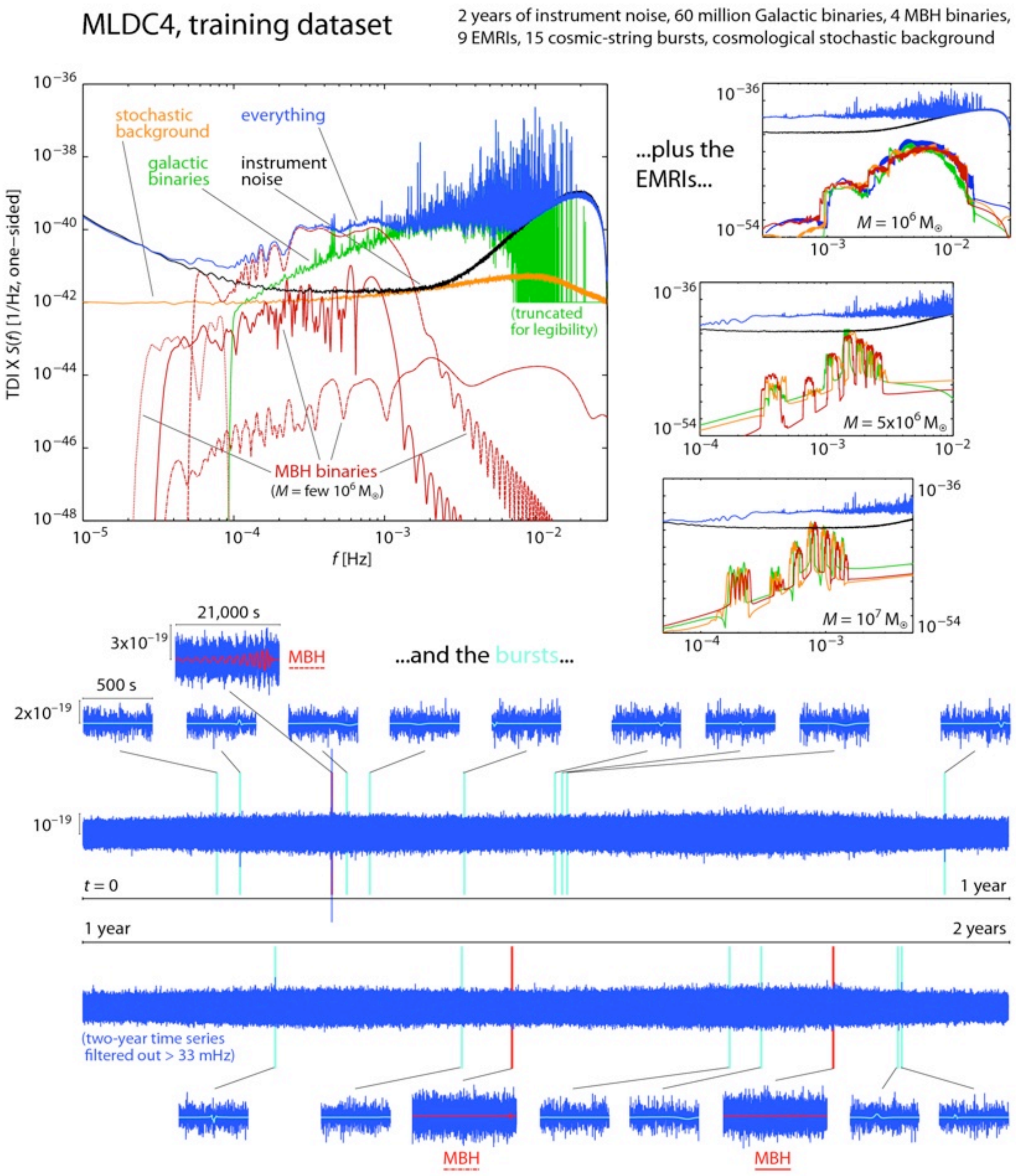
$$h_{\mathrm{GR}}(f) = A_{\mathrm{GR}}(f)e^{i\Psi_{\mathrm{GR}}(f)} \qquad \delta\Psi(f) = \beta\, u^{b}$$



at least for simple circular inspirals, stealth bias (significant $\delta\theta_{\mathrm{th}}$, non-detectable MG) is generic!

[but see arXiv/1301.2627]

the Mock LISA Data Challenges: successfully fostering and consolidating collaborative data-analysis development



- five challenges completed between 2006 and 2011
- 70 participants, 25 institutions 30+ publications
- demonstrated the detection and parameter estimation of all major LF GW source classes, using a great variety of methods
- lisatools.googlecode.com lisasolve.googlecode.com
- acknowledged model for LISA Pathfinder and the International Pulsar-Timing Array