

MACHINE LEARNING IN HIGH-ENERGY PHYSICS

BALÁZS KÉGL

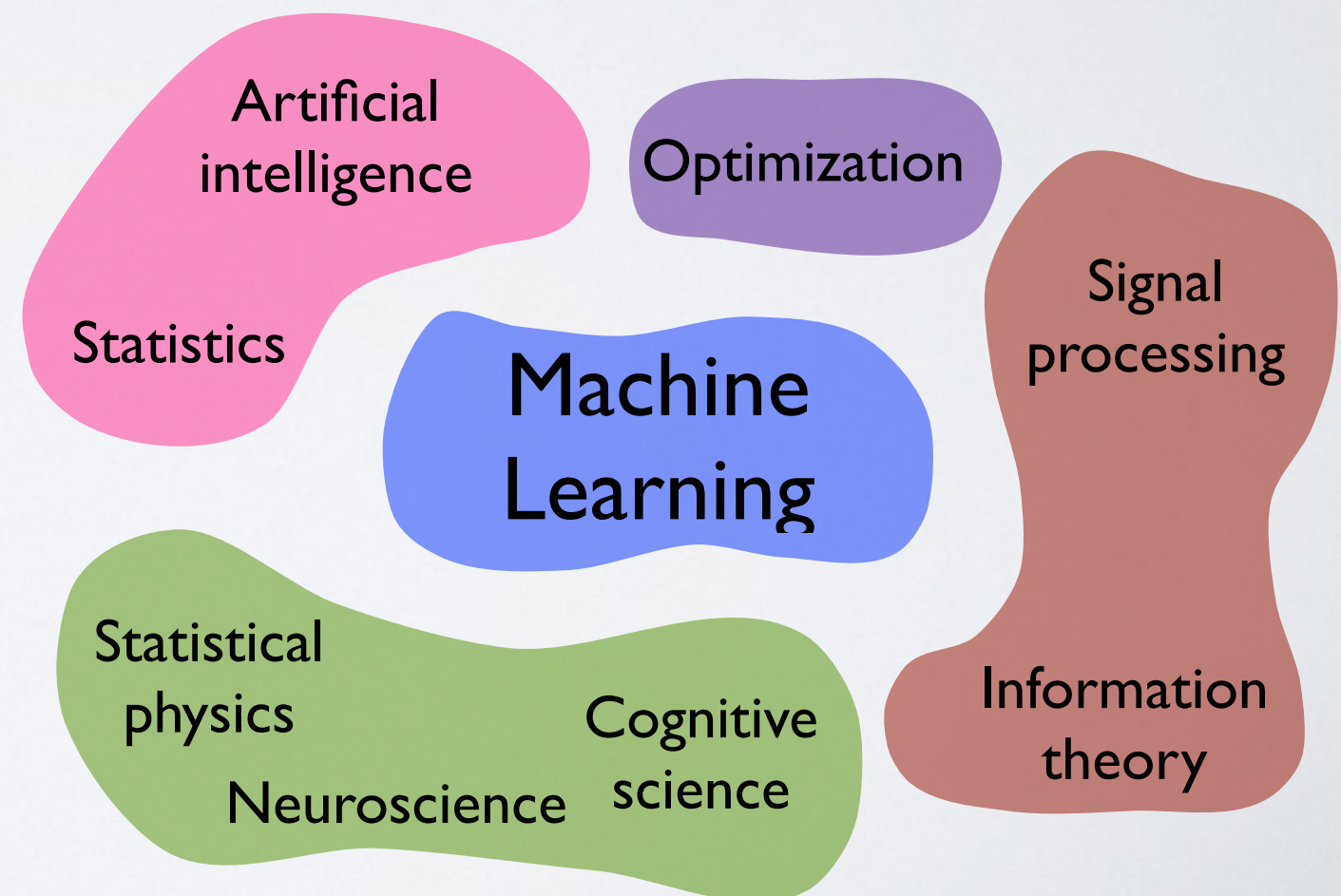
Linear Accelerator Laboratory and Computer Science Laboratory
CNRS/IN2P3 & University Paris-S{ud,acly}

CERN, September 5, 2014

OUTLINE

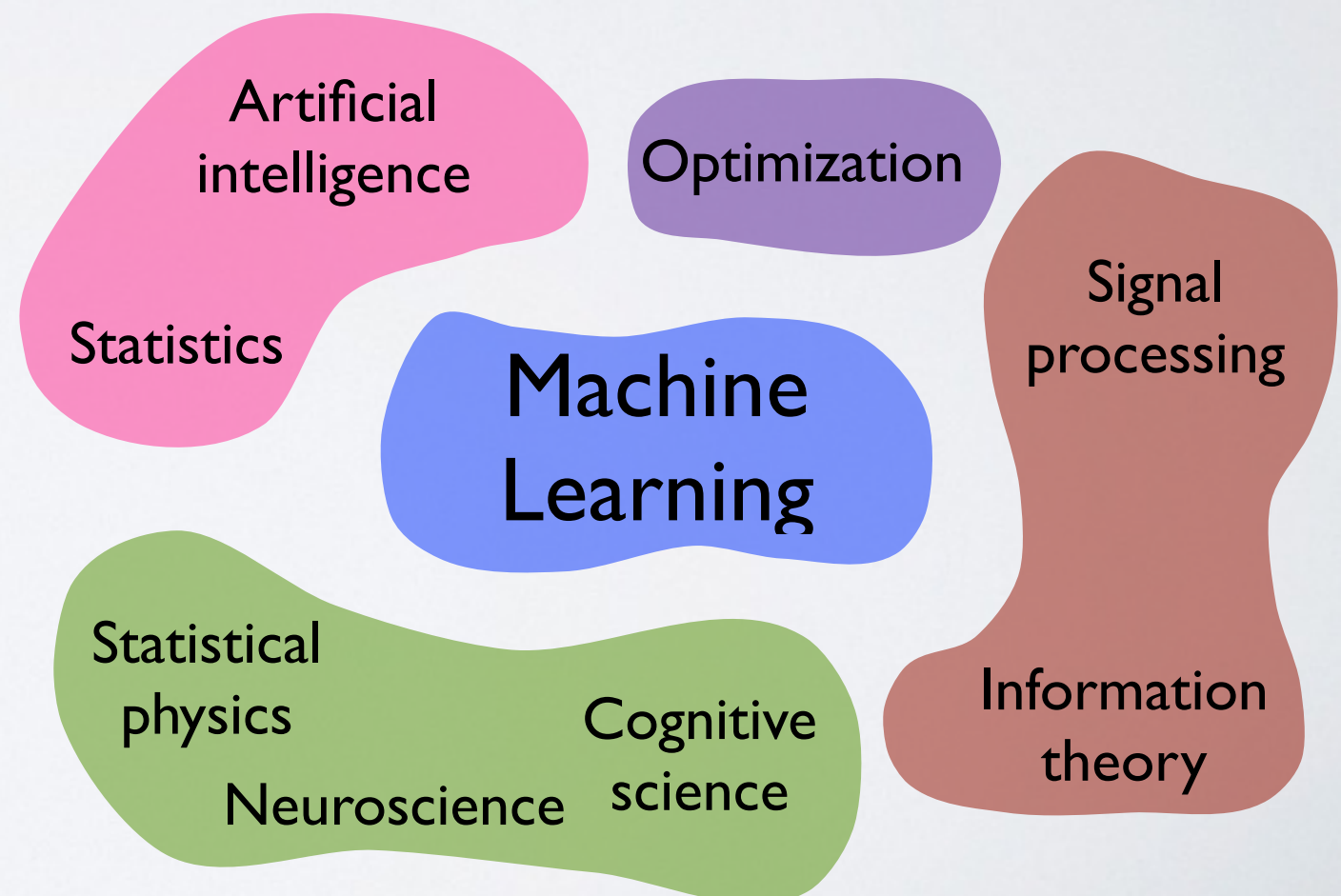
- What is machine learning/data science?
- Two projects to illustrate ML in HEP
 - budgeted learning for triggers (LHCb)
 - classification for discovery and the HiggsML challenge (ATLAS)

WHAT IS MACHINE LEARNING?



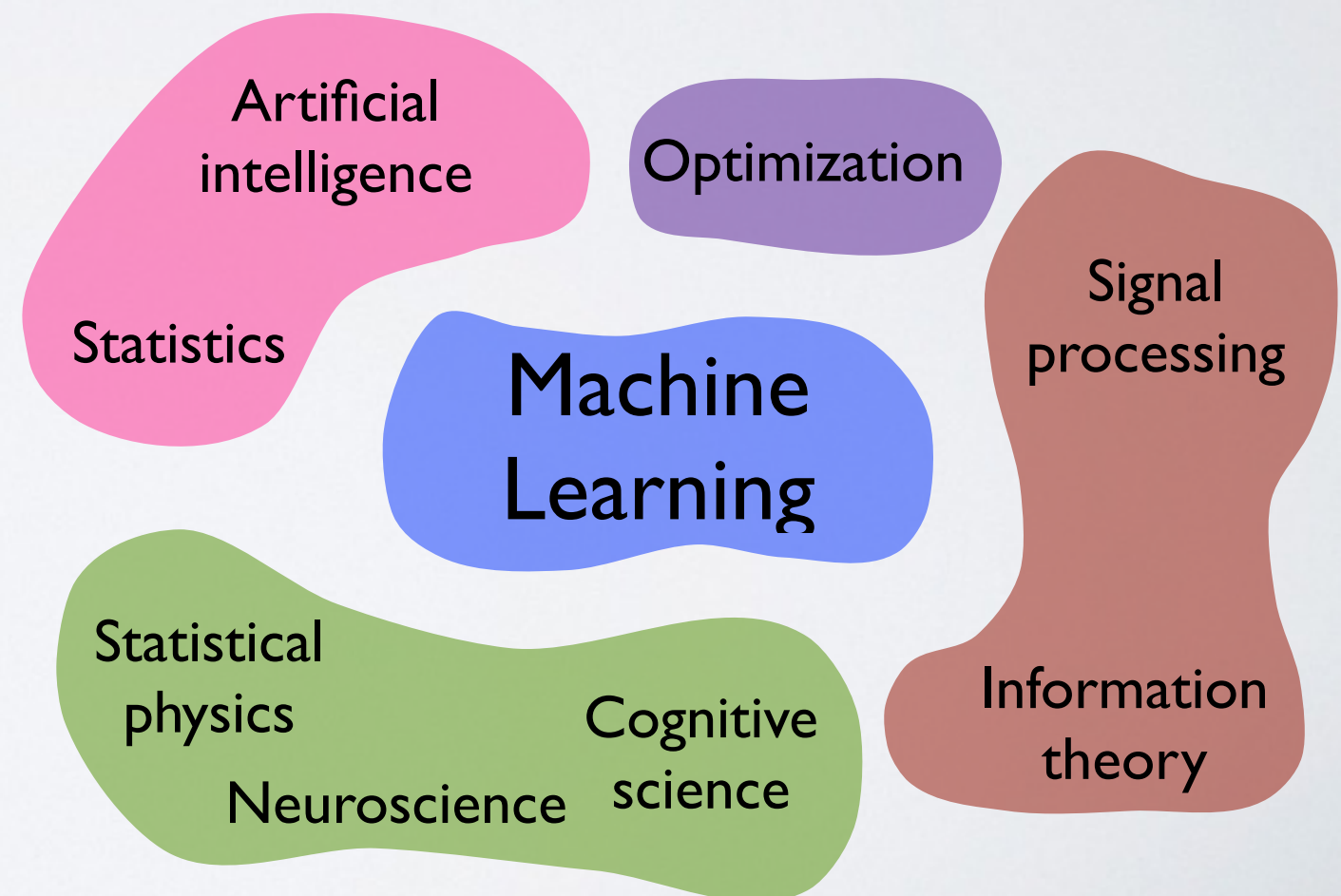
WHAT IS MACHINE LEARNING?

- “The science of getting computers to act **without being explicitly programmed**” - Andrew Ng (Stanford/Coursera)



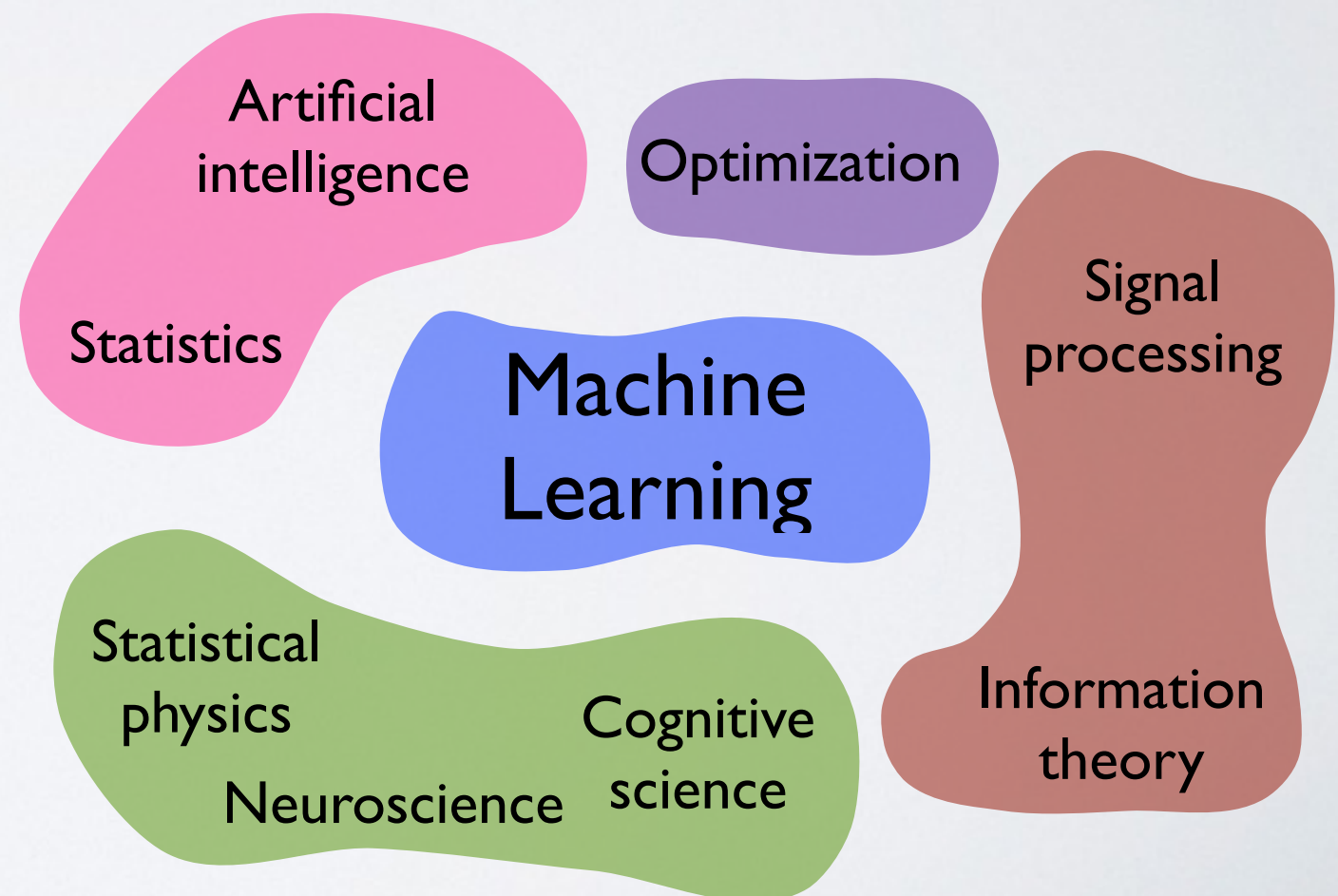
WHAT IS MACHINE LEARNING?

- “The science of getting computers to act **without being explicitly programmed**” - Andrew Ng (Stanford/Coursera)
- part of standard **computer science** curriculum since the 90s



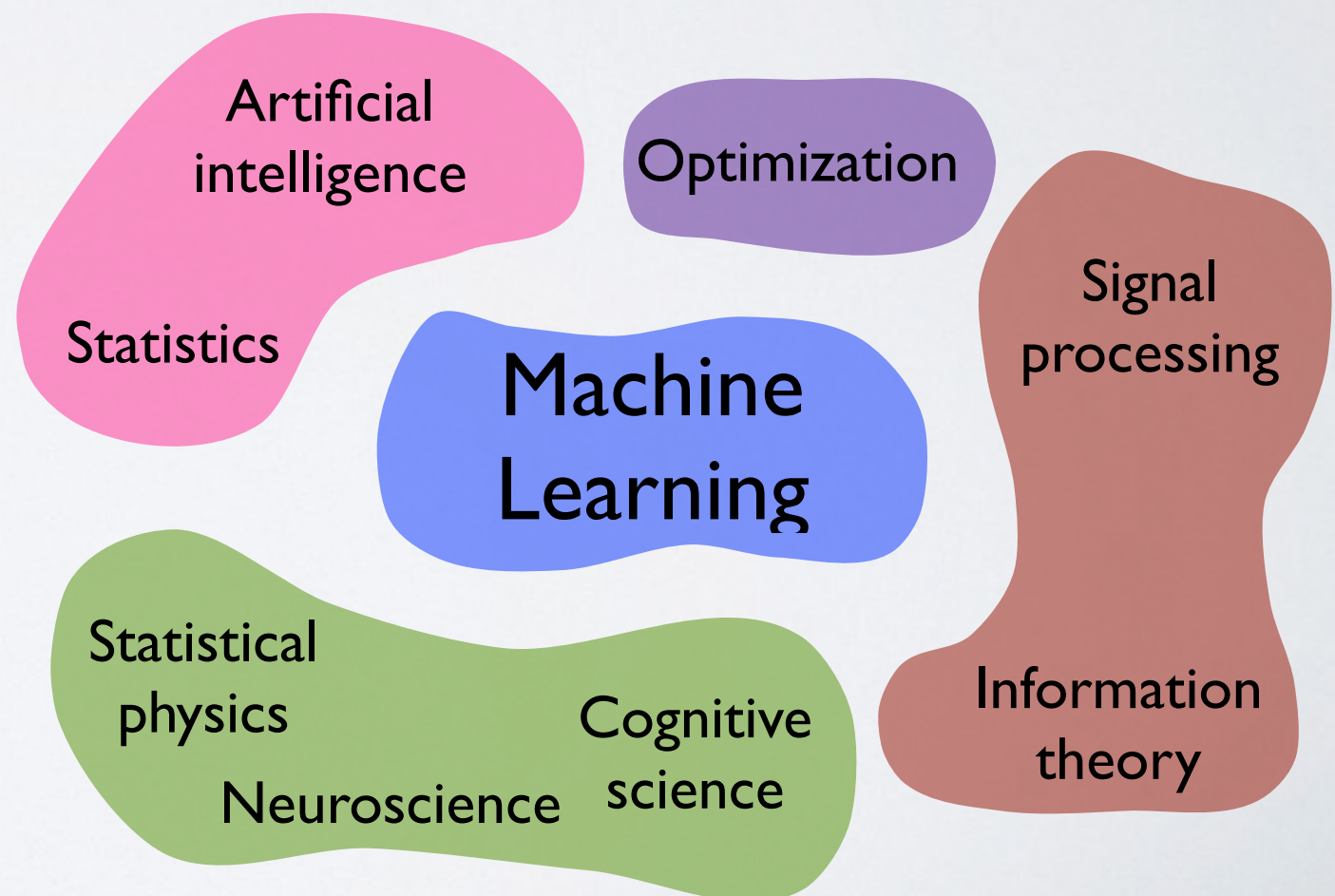
WHAT IS MACHINE LEARNING?

- “The science of getting computers to act **without being explicitly programmed**” - Andrew Ng (Stanford/Coursera)
 - part of standard **computer science** curriculum since the 90s
 - inferring **knowledge** from **data**



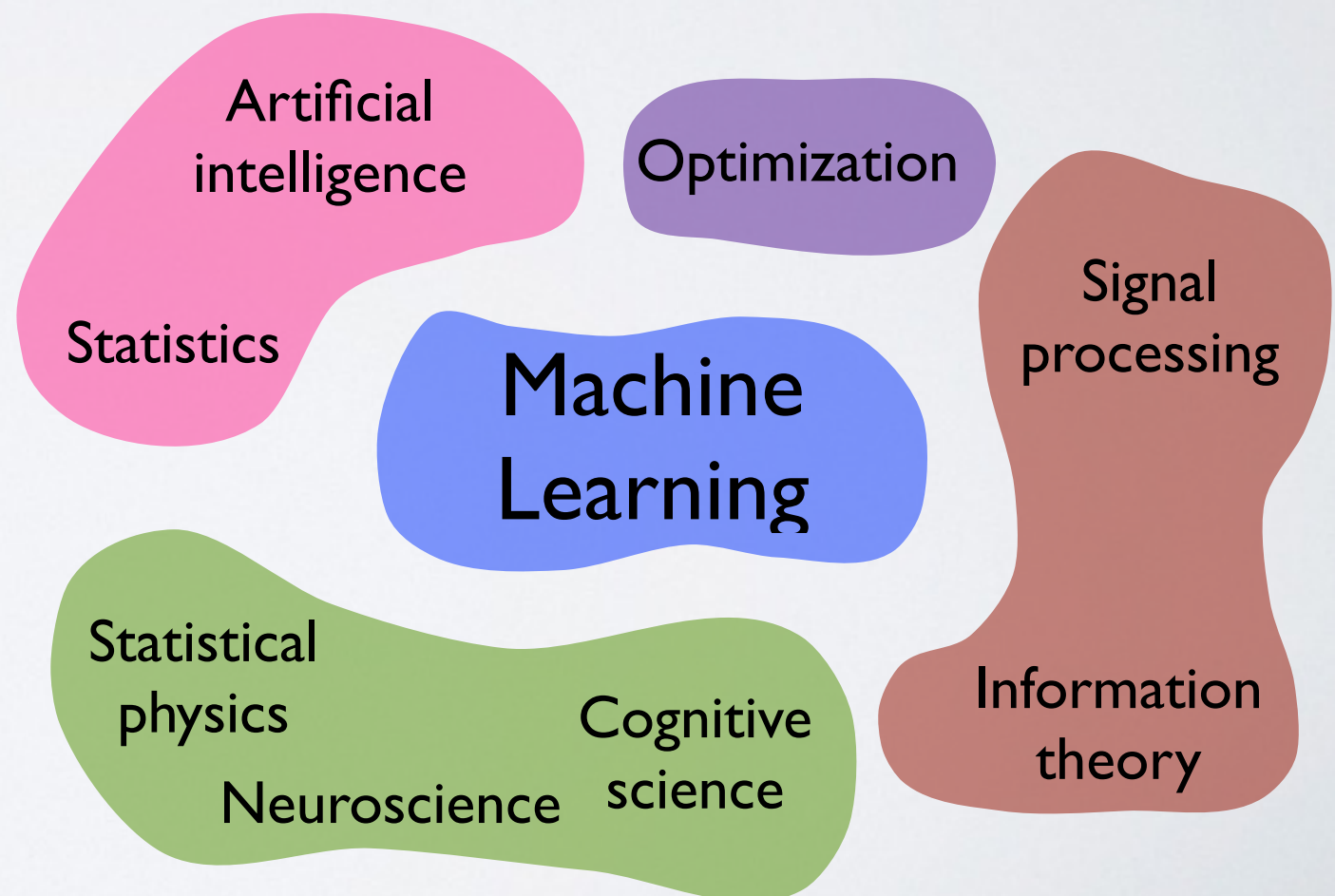
WHAT IS MACHINE LEARNING?

- “The science of getting computers to act **without being explicitly programmed**” - Andrew Ng (Stanford/Coursera)
 - part of standard **computer science** curriculum since the 90s
 - inferring **knowledge** from **data**
 - **generalizing** to **unseen** data



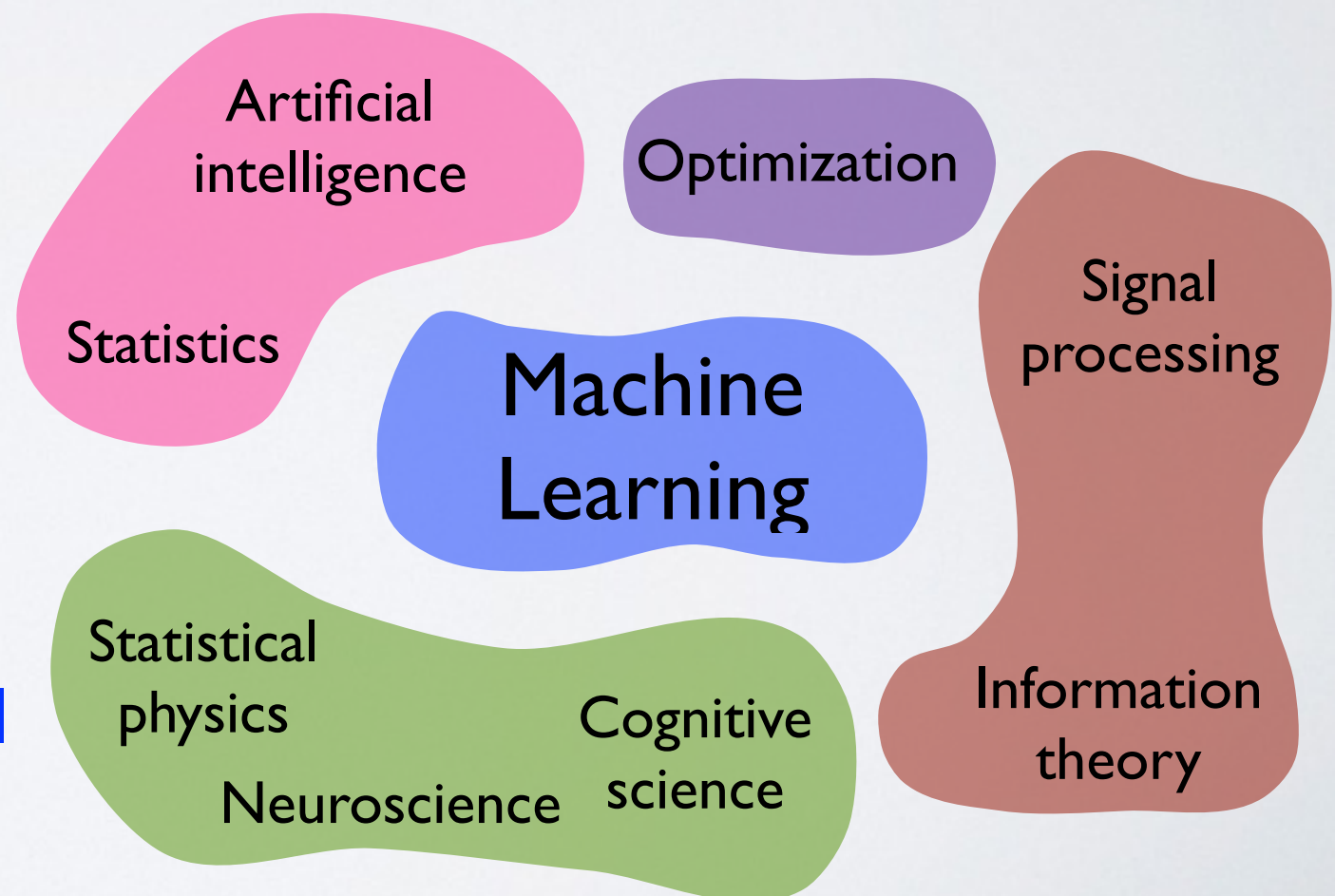
WHAT IS MACHINE LEARNING?

- “The science of getting computers to act **without being explicitly programmed**” - Andrew Ng (Stanford/Coursera)
 - part of standard **computer science** curriculum since the 90s
 - inferring **knowledge** from **data**
 - **generalizing** to **unseen** data
 - usually **no parametric model** assumptions



WHAT IS MACHINE LEARNING?

- “The science of getting computers to act **without being explicitly programmed**” - Andrew Ng (Stanford/Coursera)
 - part of standard **computer science** curriculum since the 90s
 - inferring **knowledge** from **data**
 - **generalizing** to **unseen** data
 - usually **no parametric model** assumptions
 - emphasizing the **computational challenges**



MACHINE LEARNING TAXONOMY

MACHINE LEARNING TAXONOMY

- **Supervised** learning: non-parametric (model-free) **input - output** functions
 - **classification** (Trees, BDT, SVM, NN) - what you call MVA
 - **regression** (Trees, NN, Gaussian Processes)
- **Unsupervised** learning: non-parametric **data representation**
 - **clustering** (k-means, spectral clustering, Dirichlet processes)
 - **dimensionality reduction** (PCA, ISOMAP, LLE, auto-associative NN)
 - **density estimation** (kernel density, Gaussian mixtures, the Boltzmann machine)
- **Reinforcement** learning:
 - learning + dynamic control: learn to **behave in an environment** to maximize cumulative reward

MACHINE LEARNING TAXONOMY

- **Supervised** learning: non-parametric (model-free) **input - output** functions
 - **classification** (Trees, BDT, SVM, NN) - what you call MVA
 - **regression** (Trees, NN, Gaussian Processes)
- **Unsupervised** learning: non-parametric **data representation**
 - **clustering** (k-means, spectral clustering, Dirichlet processes)
 - **dimensionality reduction** (PCA, ISOMAP, LLE, auto-associative NN)
 - **density estimation** (kernel density, Gaussian mixtures, the Boltzmann machine)
- **Reinforcement** learning:
 - learning + dynamic control: learn to **behave in an environment** to maximize cumulative reward

MACHINE LEARNING RESEARCH

We make our living by **inventing techniques**

MACHINE LEARNING RESEARCH

We are interested in **problems** at the
edges of our current methodological
capacities

CLASSIFICATION



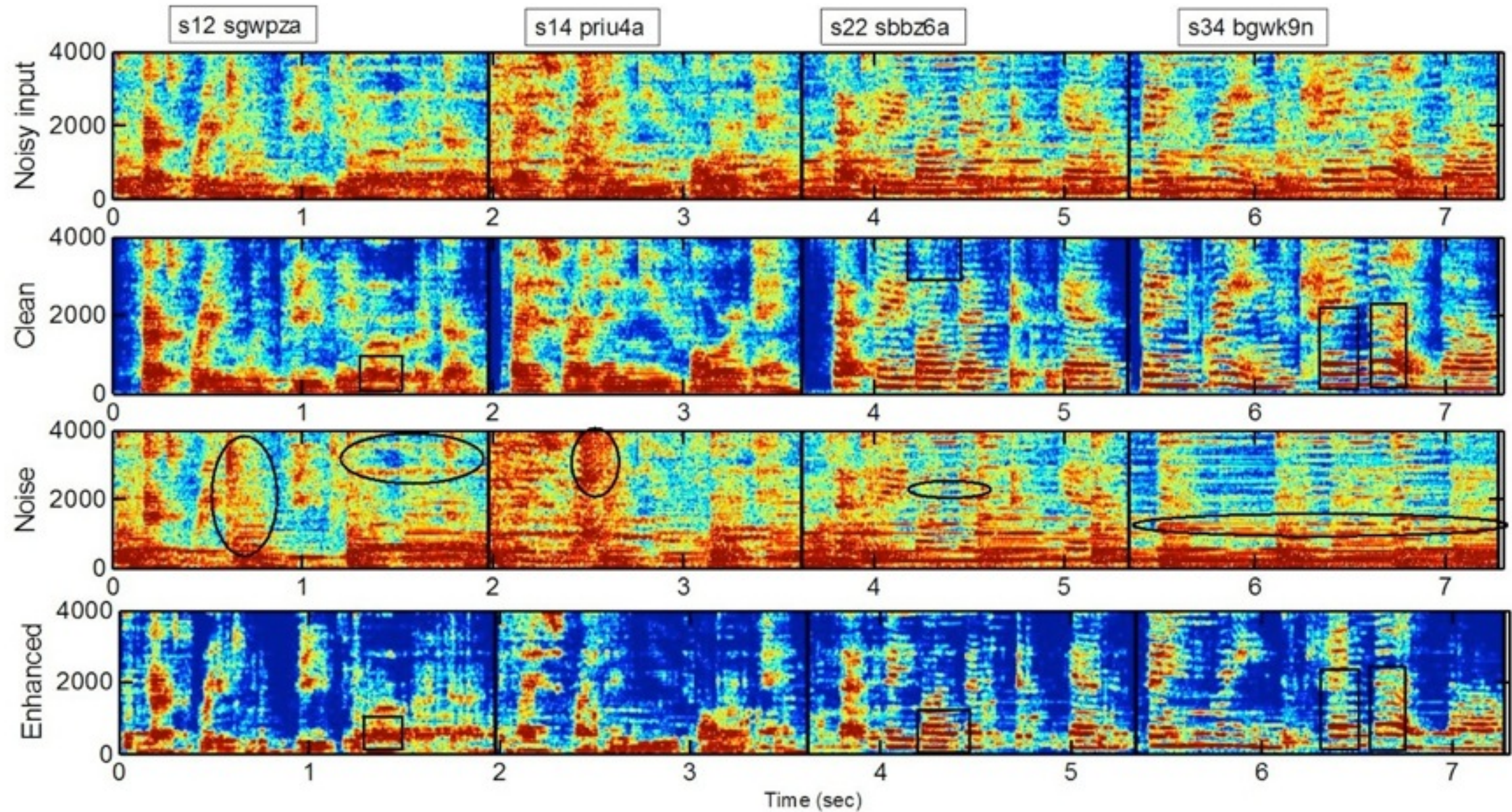
Character recognition

CLASSIFICATION



Emotion recognition

CLASSIFICATION



Speech recognition

CLASSIFICATION

- **Input**: a usually **high dimensional vector** x
- **Output**: a **category** (aka label, class) y
- Usually **no parametric model**
 - the classification function $y = g(x)$ is **learned** using a **training set**
 $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$
- **Well-tested algorithms**:
 - **neural networks**, **support vector machines**, **boosting (BDTs)**

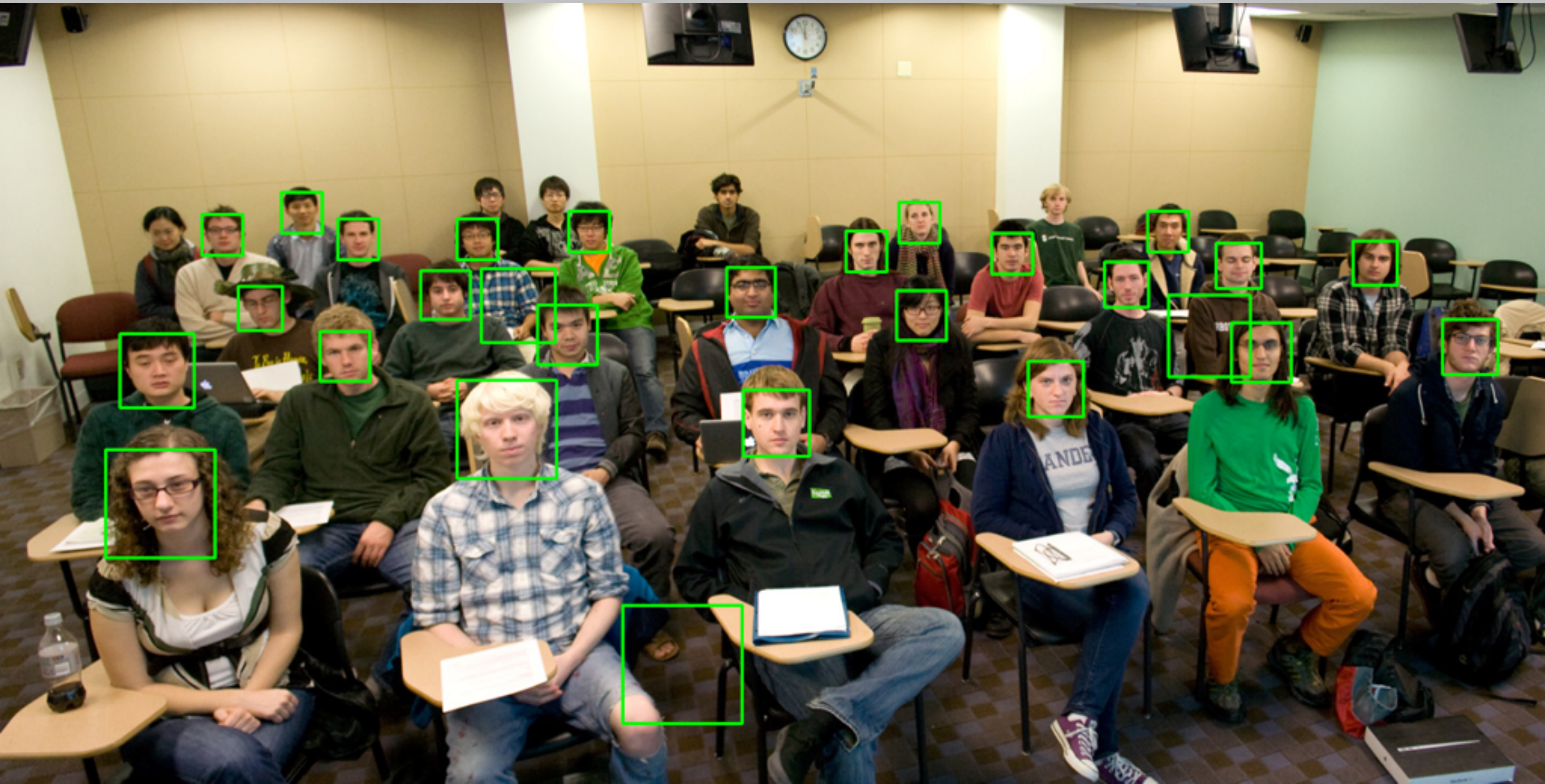
CLASSIFICATION

The only goal is a **low probability of error**

$$P(g(x) \neq y)$$

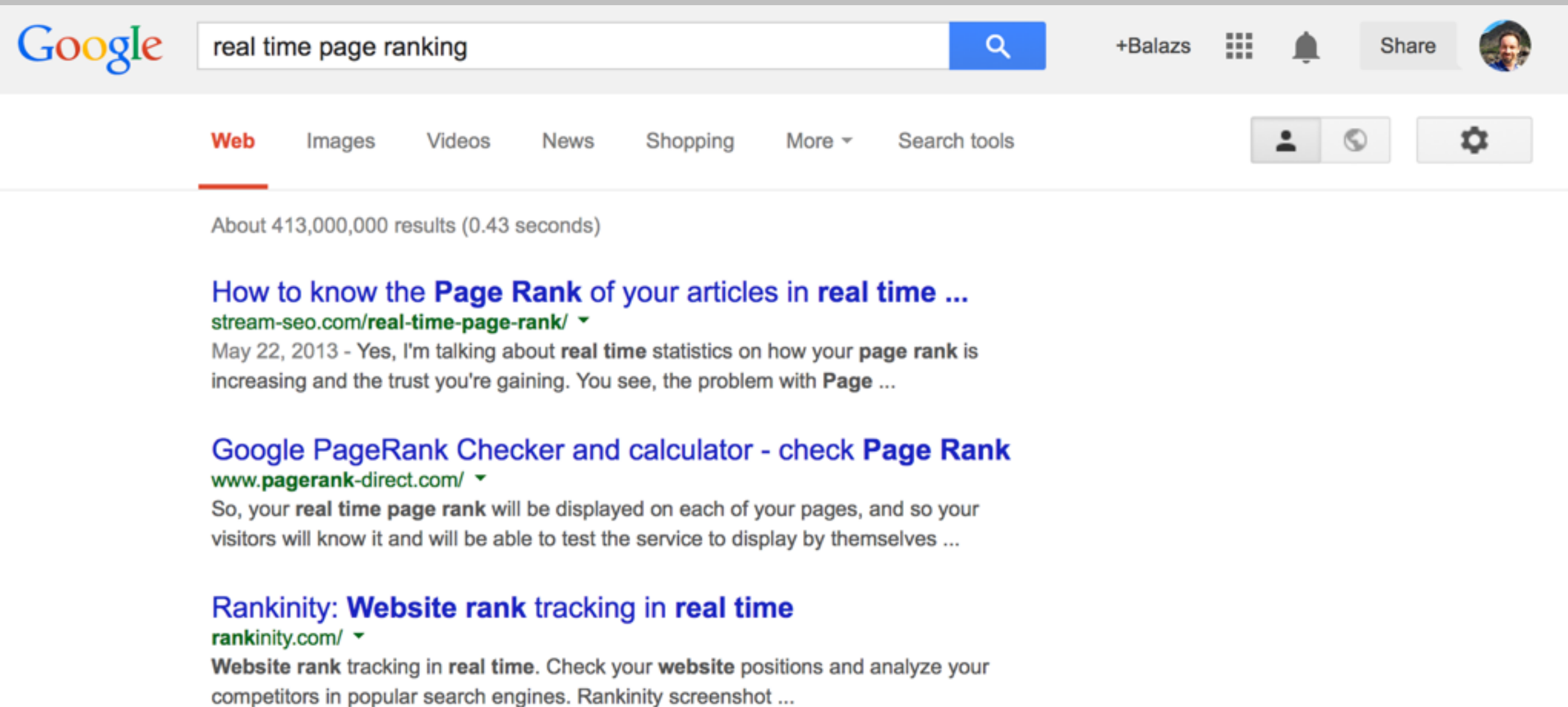
on **previously unseen** examples (x, y)

BUDGETED CLASSIFICATION



Real time face detection

BUDGETED CLASSIFICATION



The image shows a Google search results page for the query "real time page ranking". The search bar at the top contains the text "real time page ranking" and a blue search button. To the right of the search bar are links for "+Balazs", a grid icon, a bell icon, and a "Share" button with a profile picture. Below the search bar are navigation tabs for "Web", "Images", "Videos", "News", "Shopping", "More", and "Search tools". The "Web" tab is selected and underlined. Below the navigation tabs, the search results are displayed. The first result is titled "How to know the Page Rank of your articles in real time ..." from "stream-seo.com/real-time-page-rank/". The second result is titled "Google PageRank Checker and calculator - check Page Rank" from "www.pagerank-direct.com/". The third result is titled "Rankinity: Website rank tracking in real time" from "rankinity.com/".

Google

real time page ranking

+Balazs

Share

Web Images Videos News Shopping More Search tools

About 413,000,000 results (0.43 seconds)

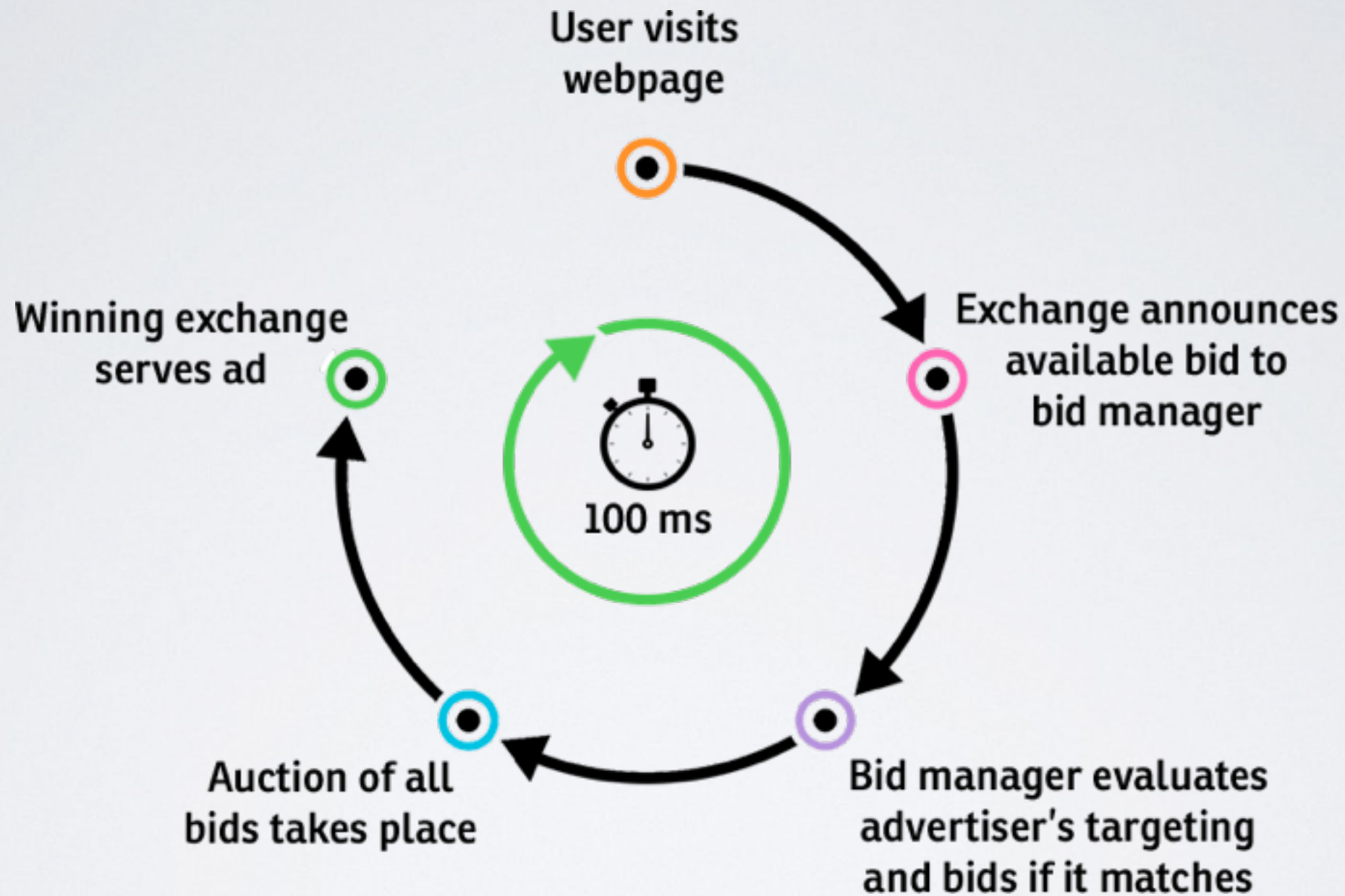
How to know the Page Rank of your articles in real time ...
stream-seo.com/real-time-page-rank/
May 22, 2013 - Yes, I'm talking about **real time** statistics on how your **page rank** is increasing and the trust you're gaining. You see, the problem with **Page** ...

Google PageRank Checker and calculator - check Page Rank
www.pagerank-direct.com/
So, your **real time page rank** will be displayed on each of your pages, and so your visitors will know it and will be able to test the service to display by themselves ...

Rankinity: Website rank tracking in real time
rankinity.com/
Website rank tracking in real time. Check your **website** positions and analyze your competitors in popular search engines. Rankinity screenshot ...

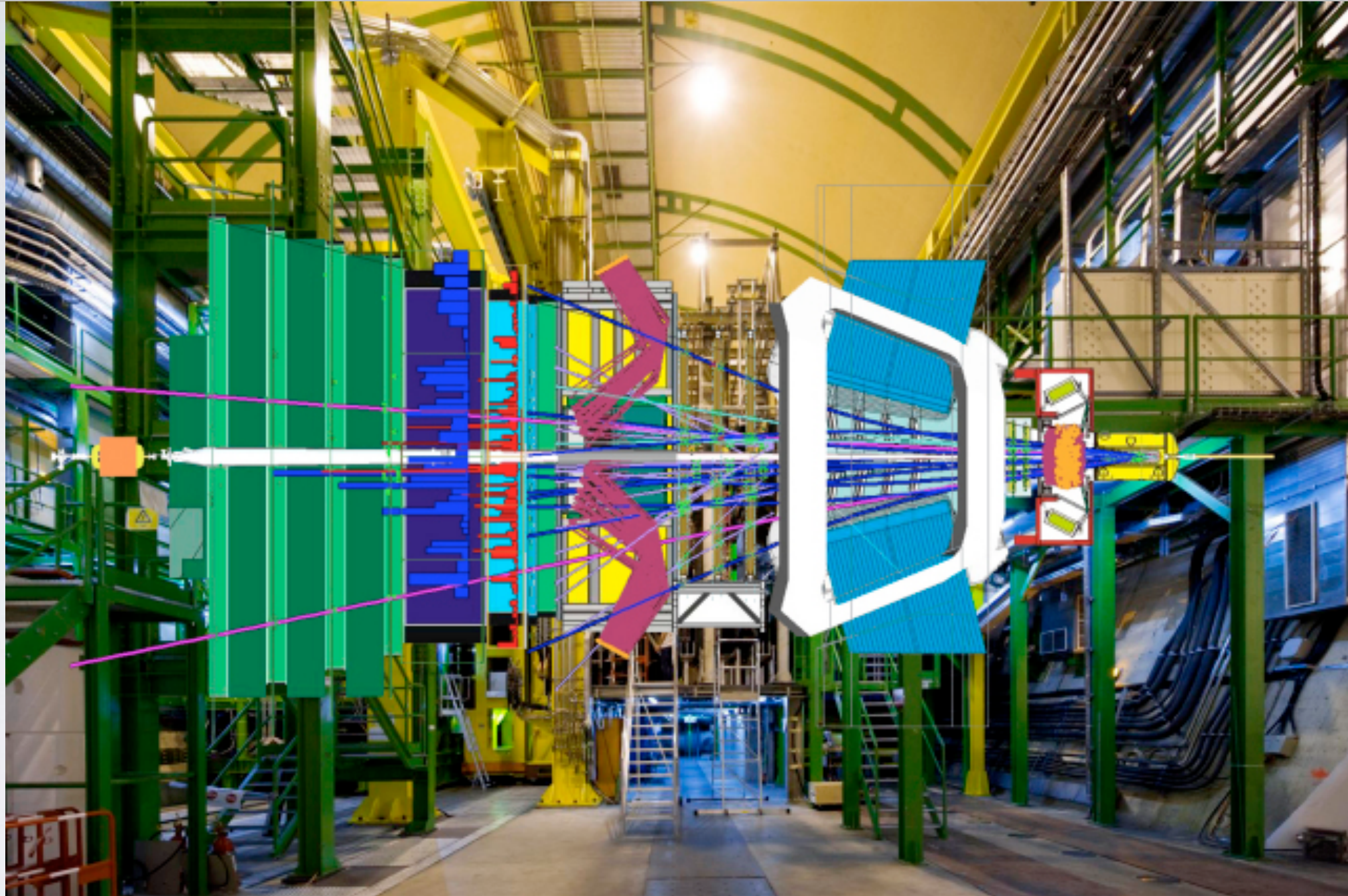
Real time web page ranking

BUDGETED CLASSIFICATION



Real time ad placement

BUDGETED CLASSIFICATION



Real time signal/background separation

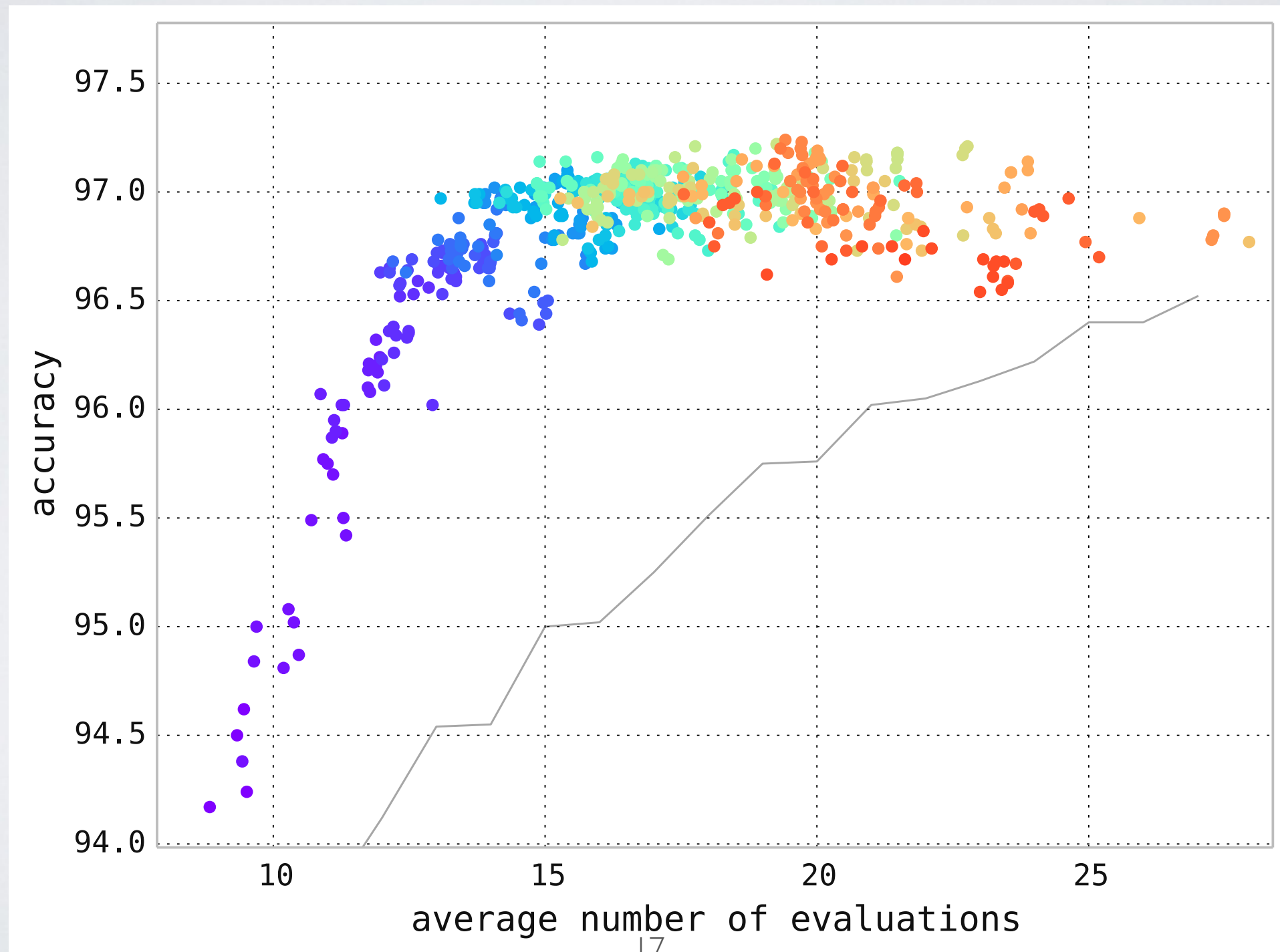
BUDGETED CLASSIFICATION

The second goal is the **fast execution** of

$g(x)$

BUDGETED CLASSIFICATION

Trade-off between quality and speed



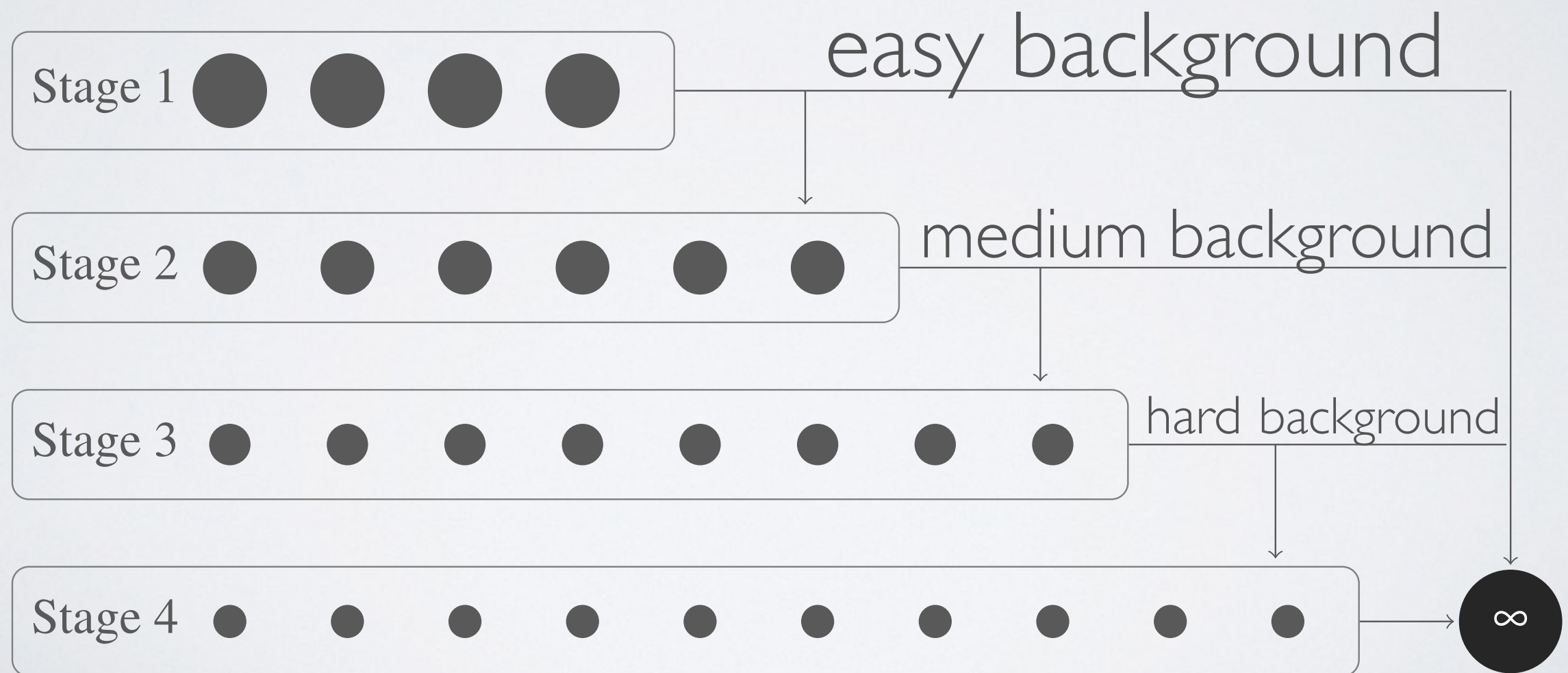
BUDGETED CLASSIFICATION

- **Time** constraints
- **Memory** constraints
- **Consumption** constraints
- **Communication** constraints

BUDGETED CLASSIFICATION

The common design:

cascade classification = trigger with levels



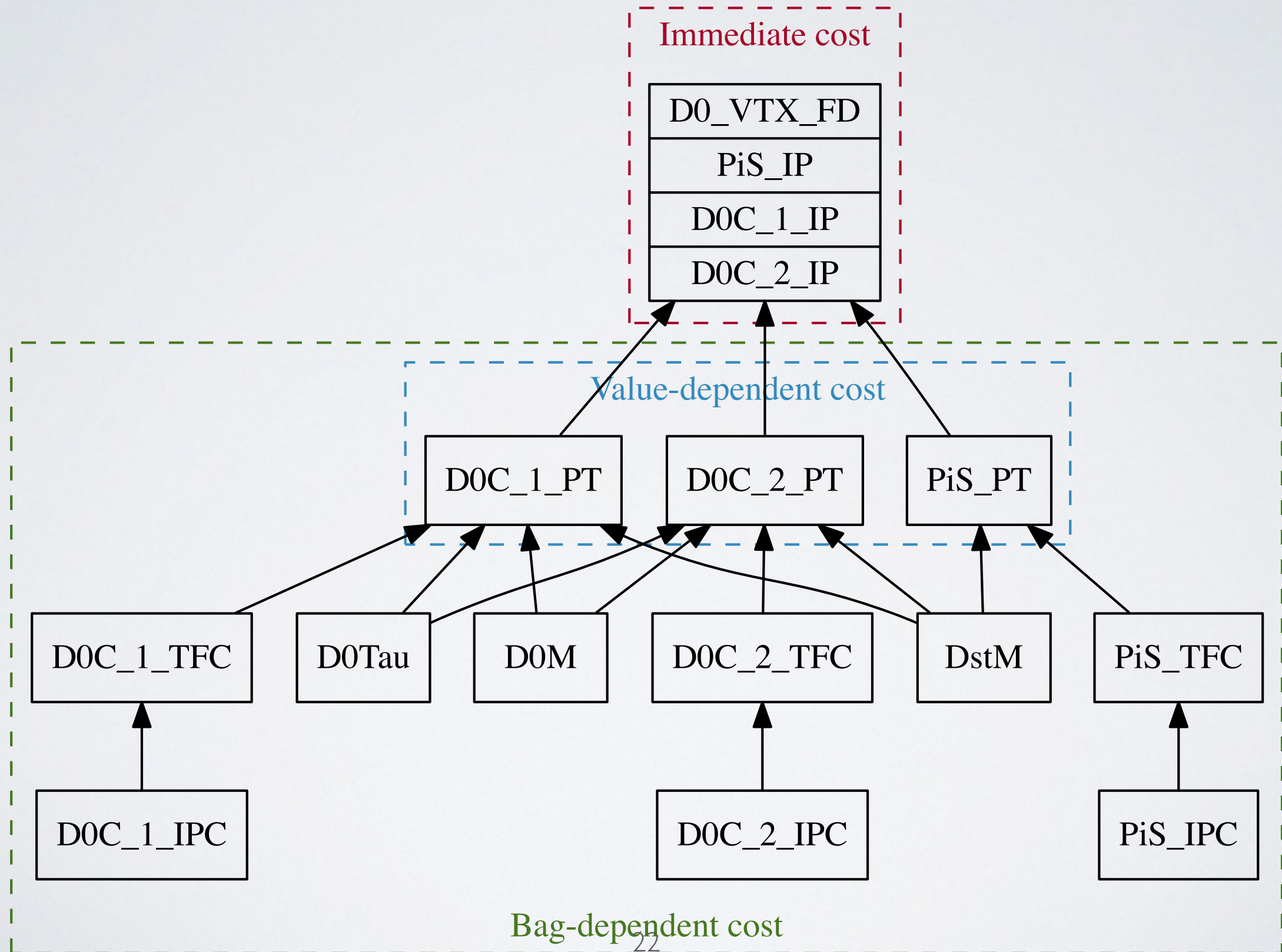
THE LHCb TRIGGER

- Collaboration with
 - Vava Gligorov (CERN)
 - Mike Williams (MIT)
 - Djalel Benbouzid (LAL)

THE LHCb TRIGGER

- A beautifully complex problem
 - **varying** feature costs
 - cost may **depend on the value**
 - events are bags of **overlapping candidates**

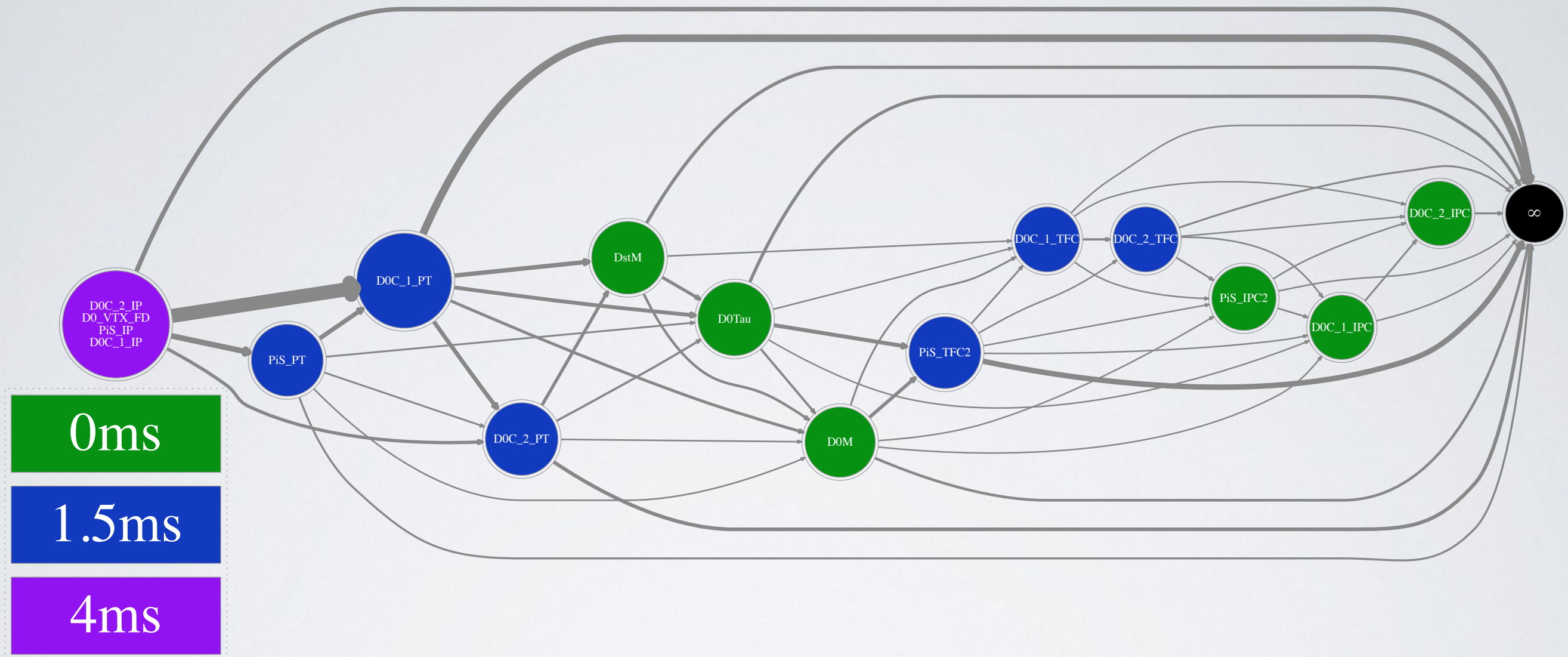
THE LHCb TRIGGER



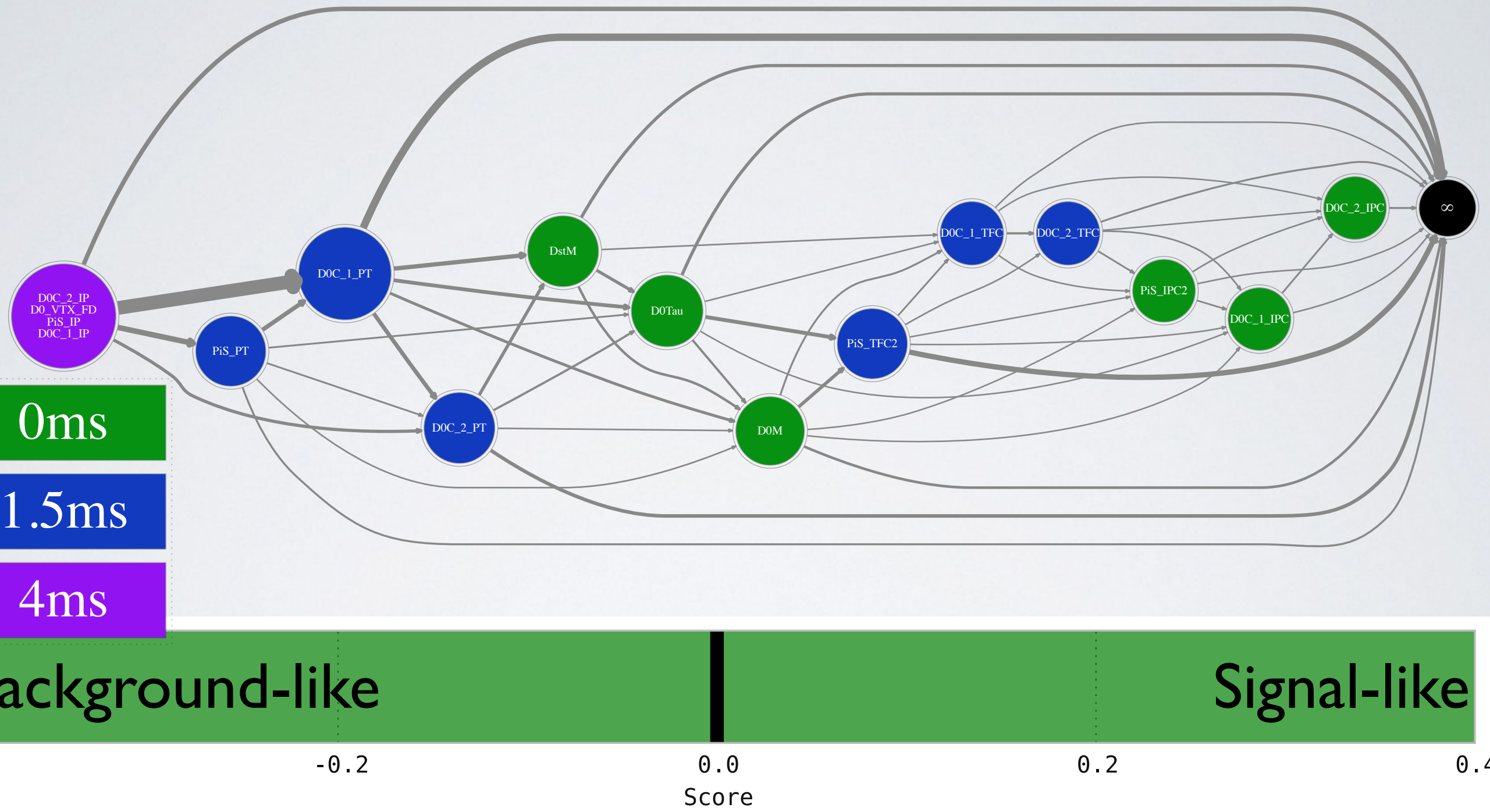
MDDAG: A SIGNAL/BCKG DECISION GRAPH

MDDAG: A SIGNAL/BCKG DECISION GRAPH

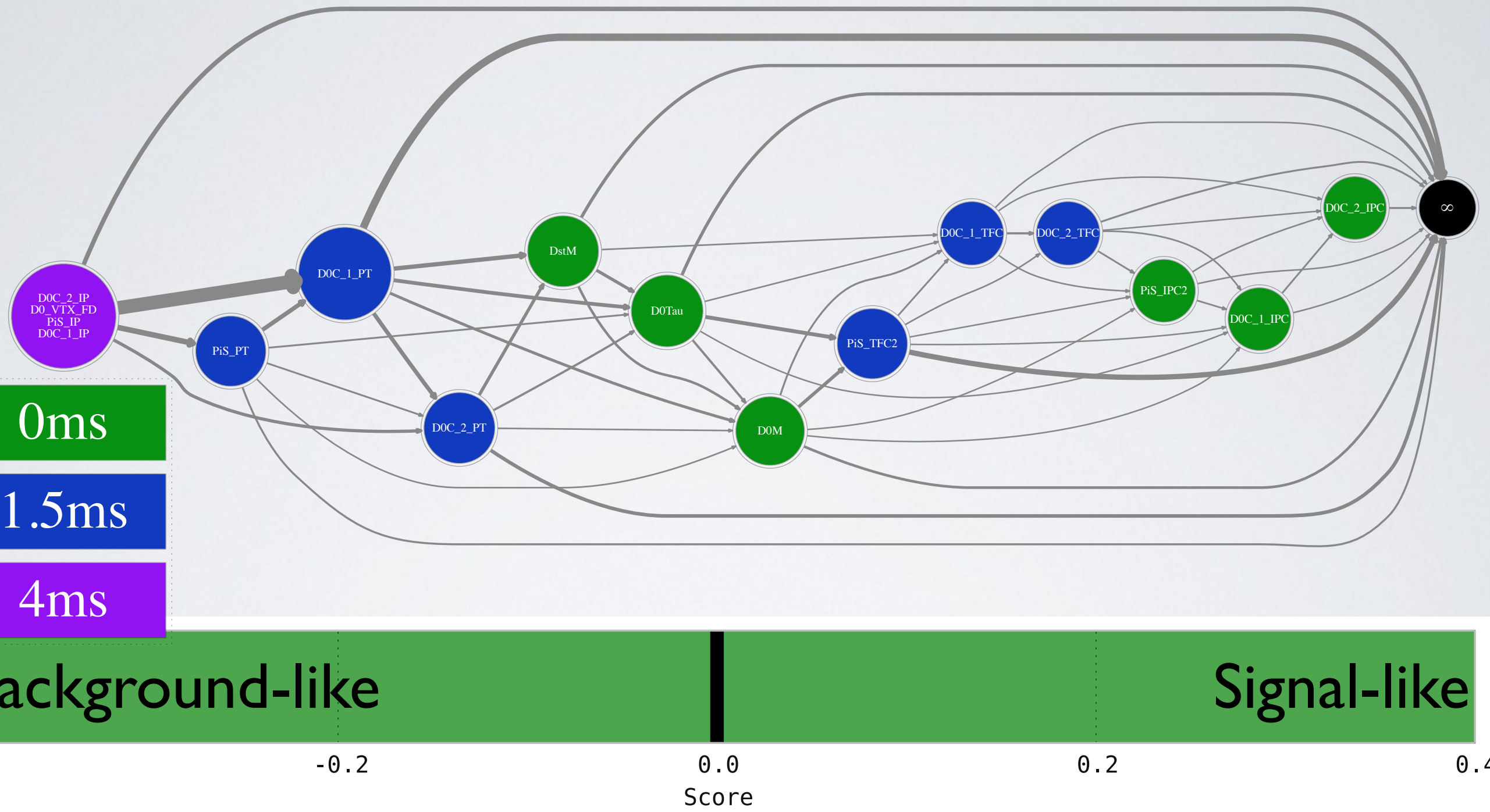
MDDAG: A SIGNAL/BCKGG DECISION GRAPH



MDDAG: A SIGNAL/BCKG DECISION GRAPH

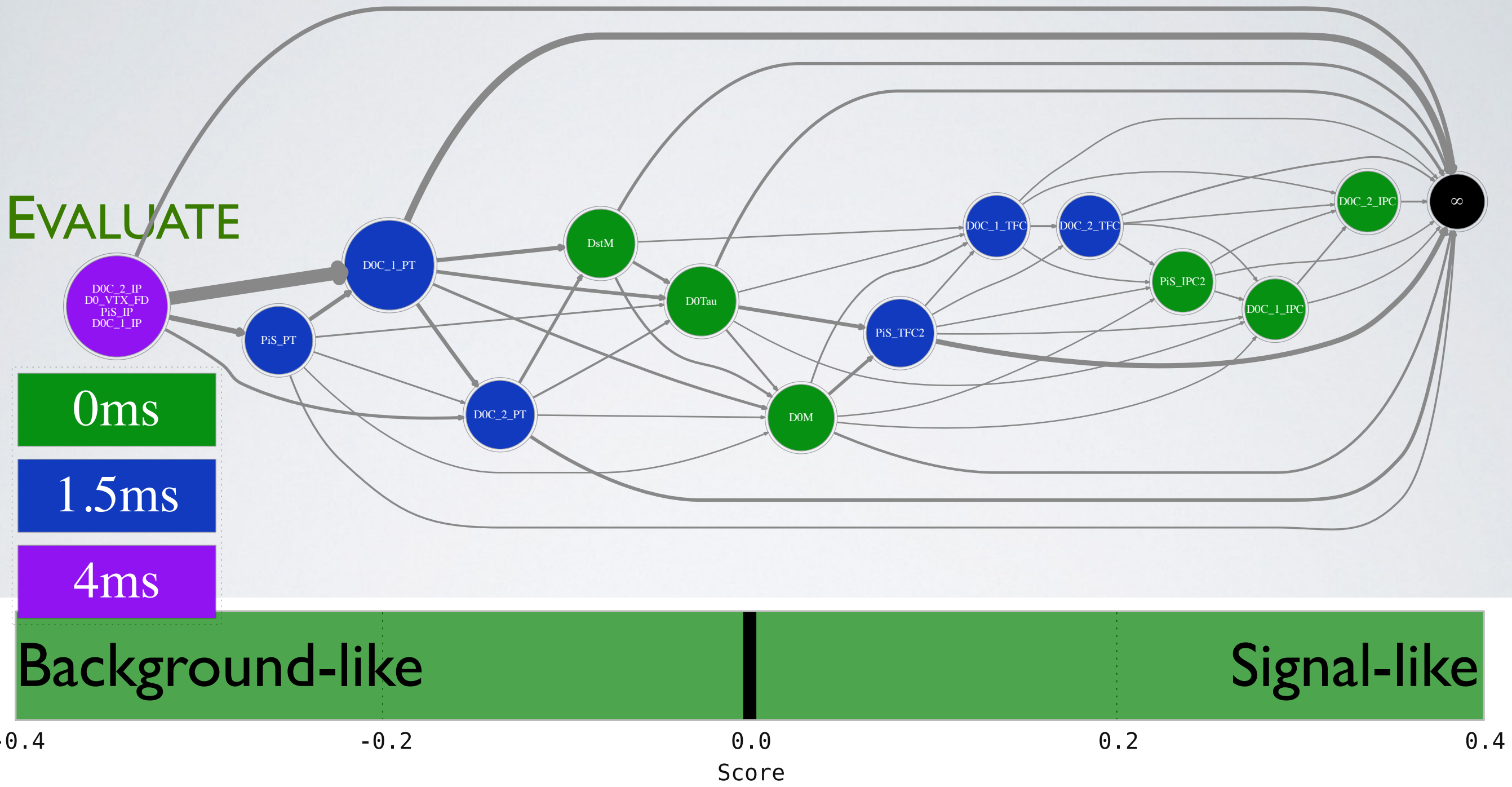


MDDAG: A SIGNAL/BCKGG DECISION GRAPH



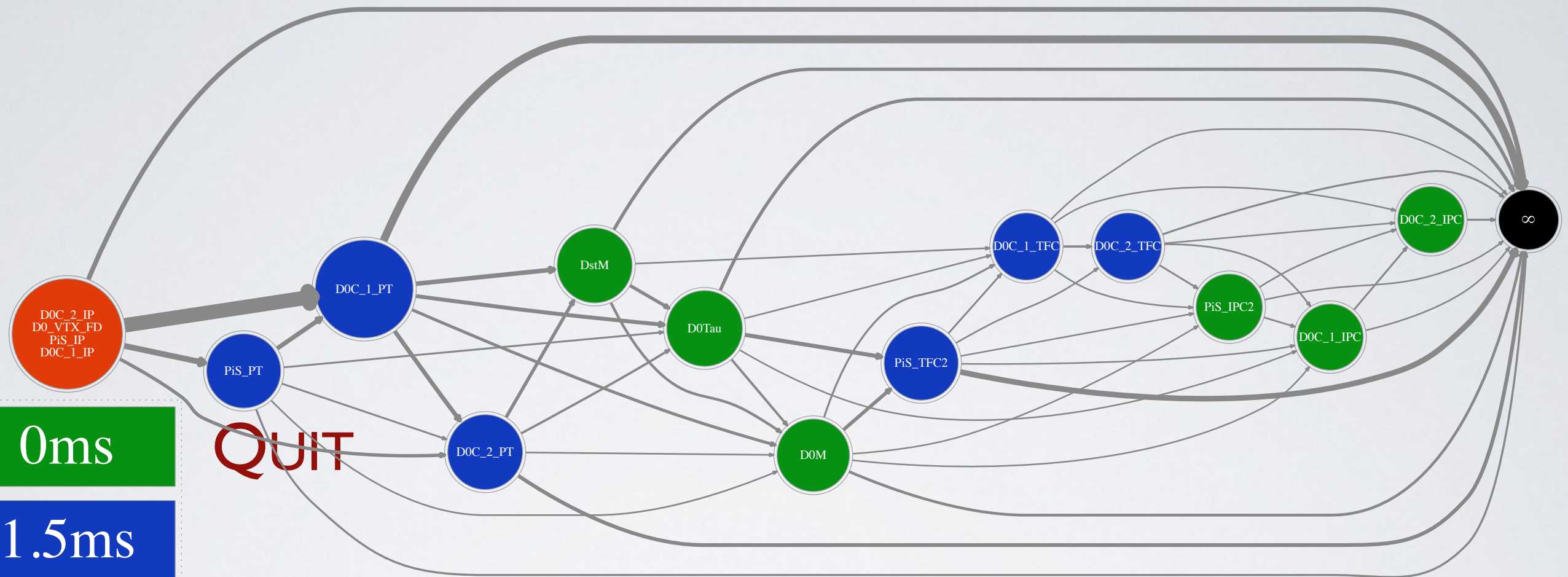
Easy background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



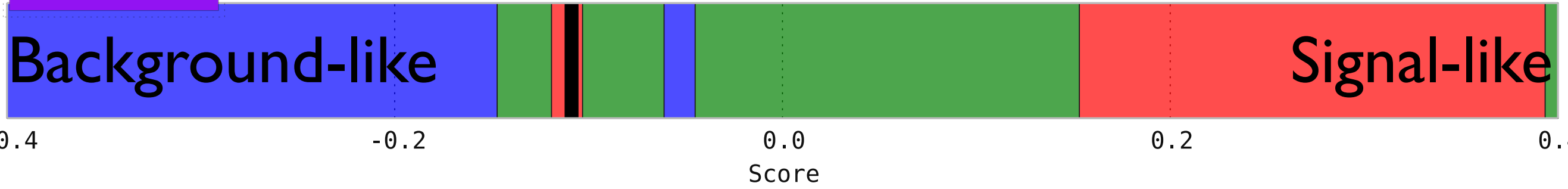
Easy background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



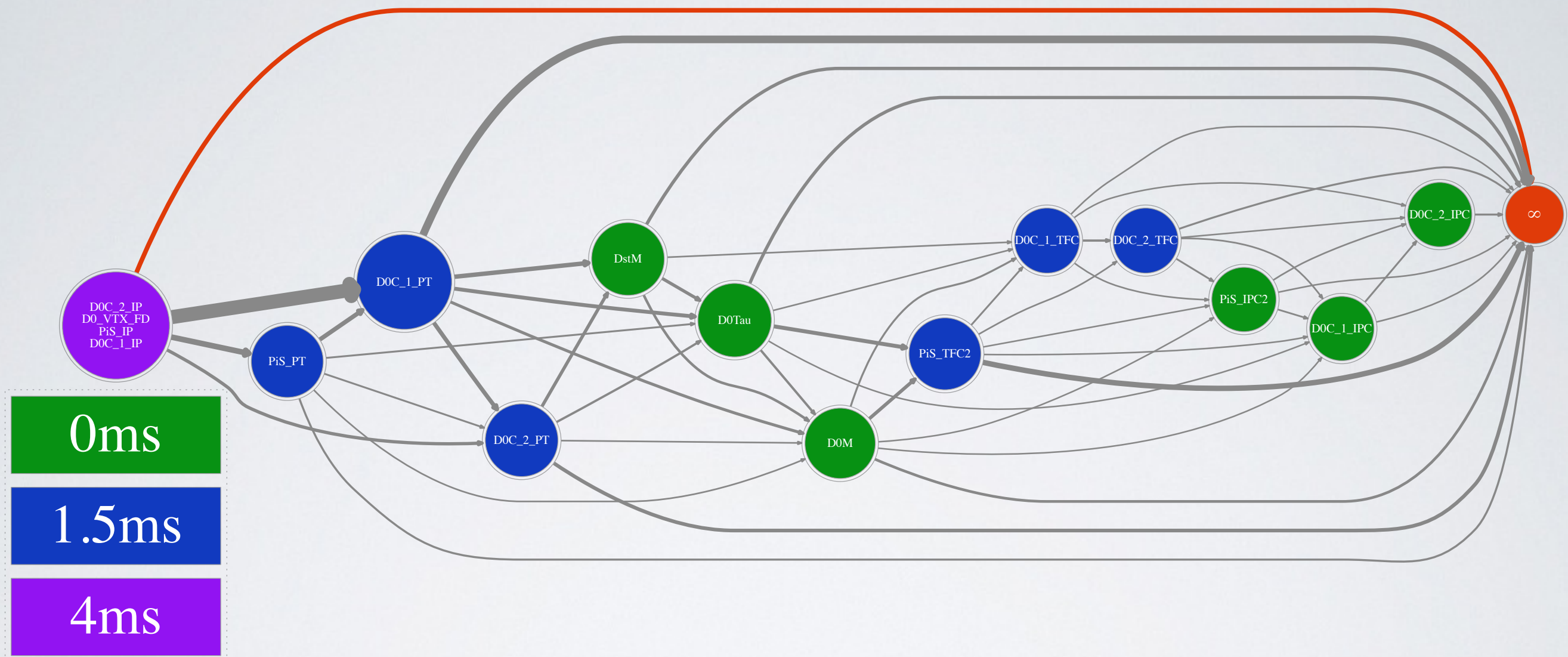
0ms
1.5ms
4ms

QUIT



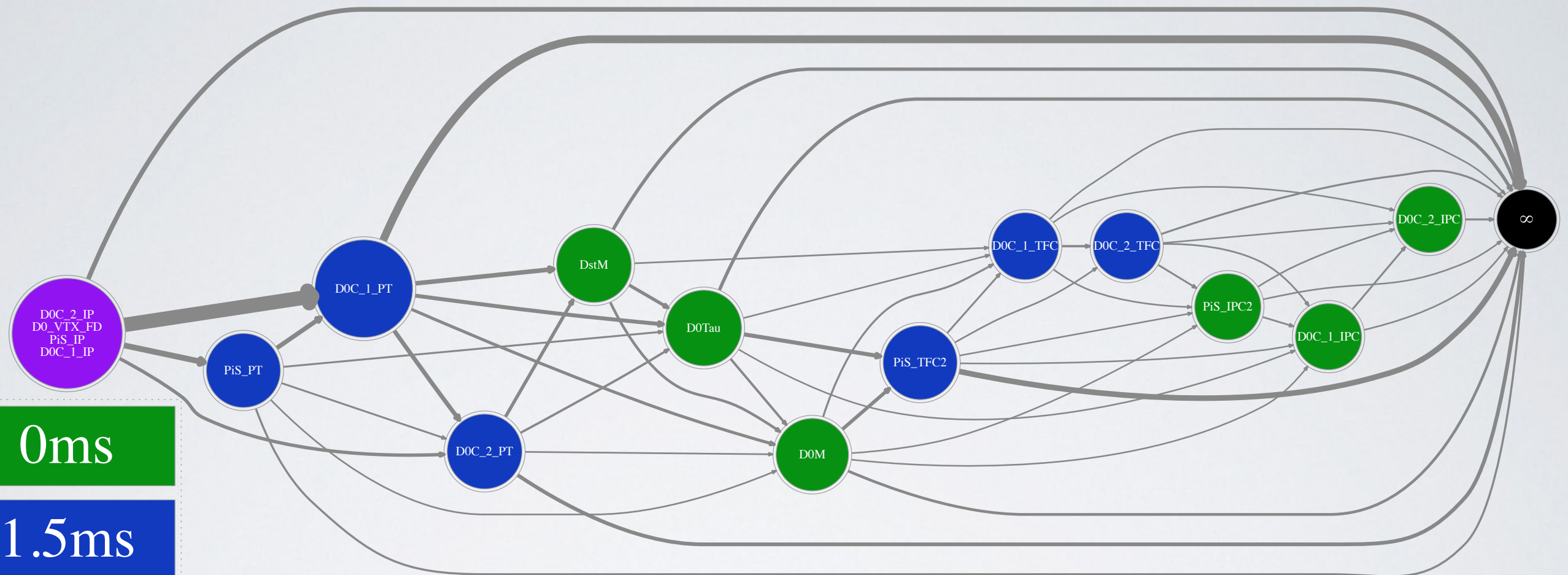
Easy background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



Easy background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



0ms

1.5ms

4ms

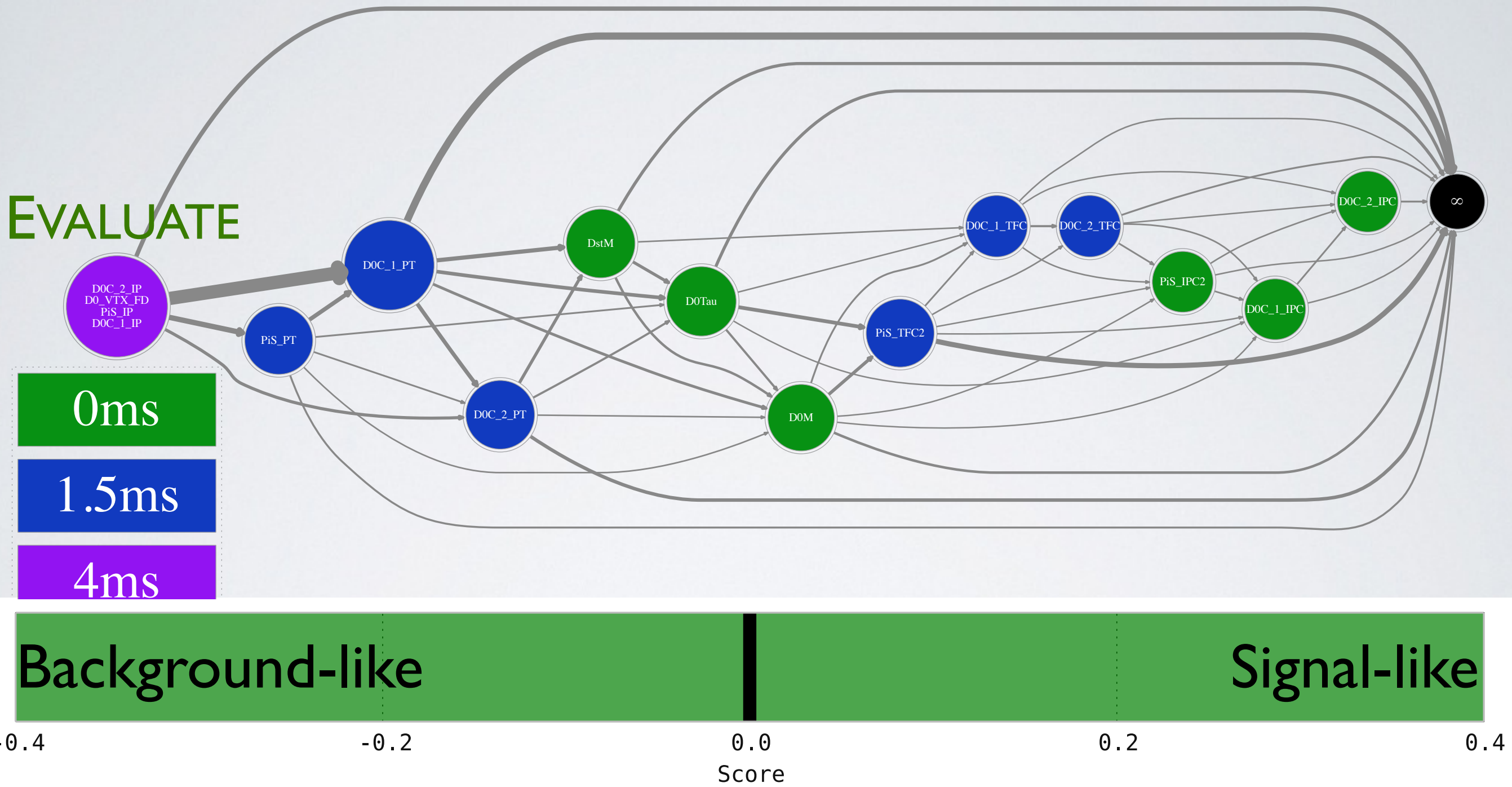
Background-like

Signal-like

Score

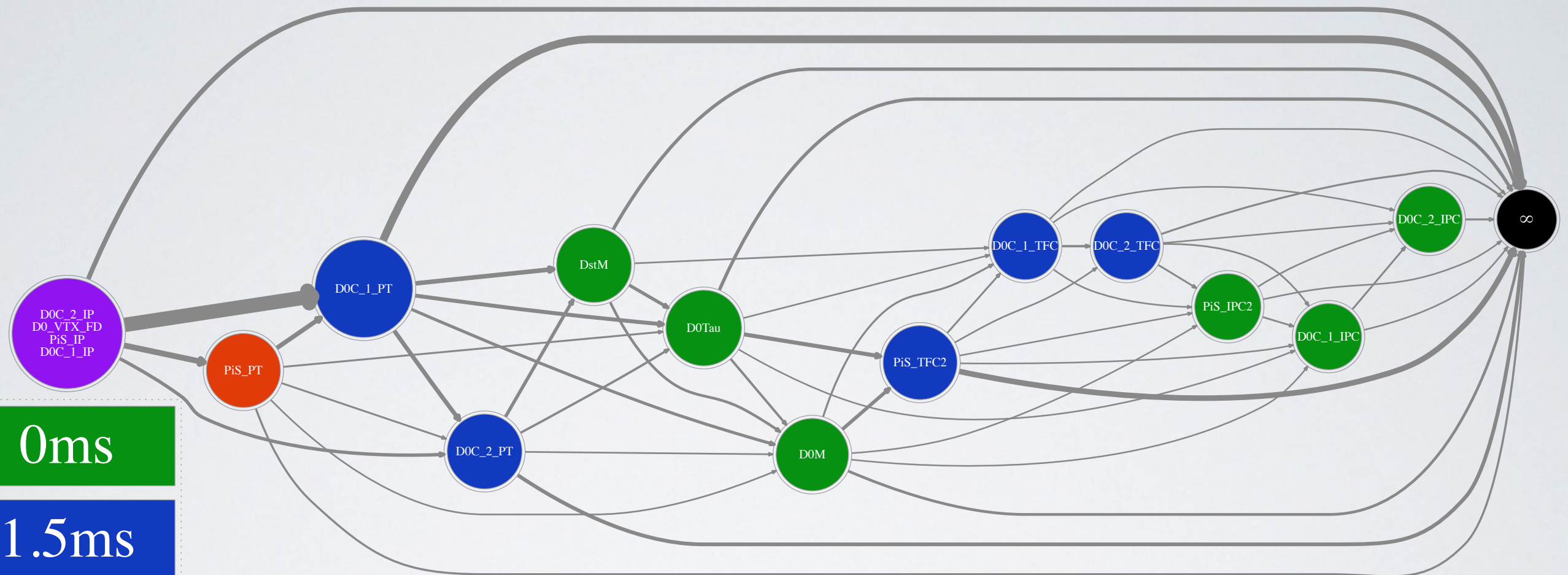
Hard background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



Hard background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



0ms

1.5ms

4ms

Background-like

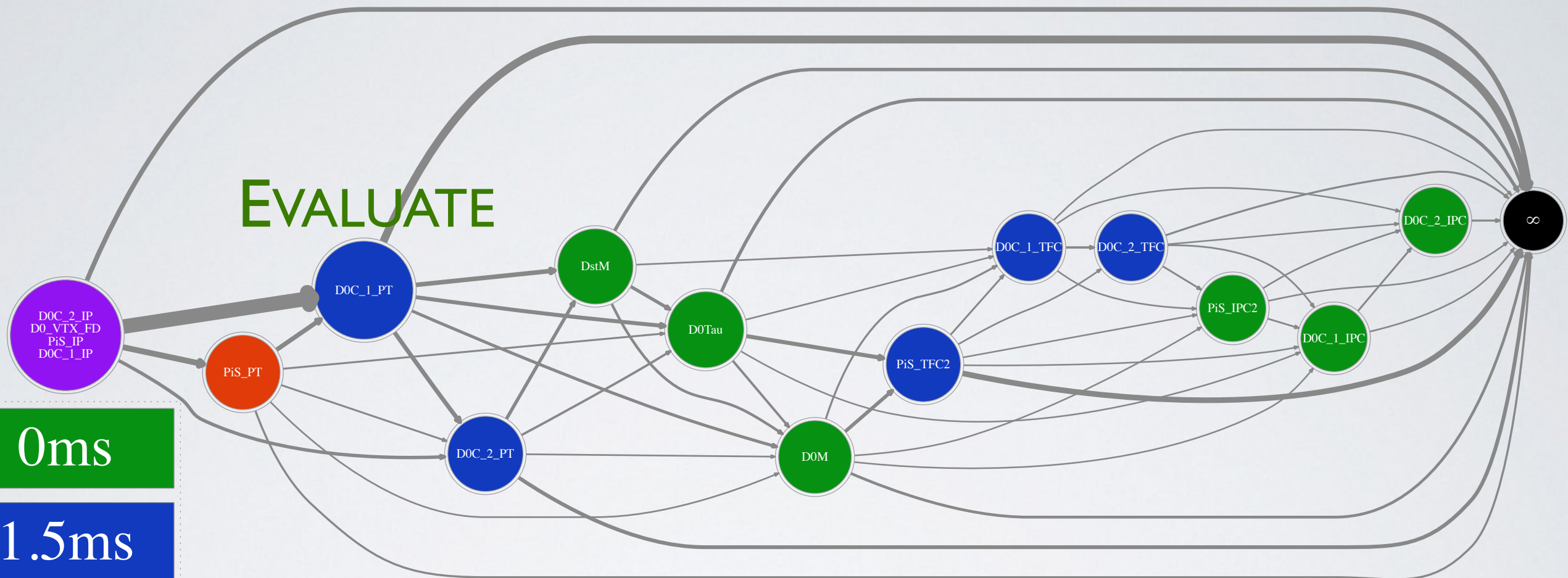
Signal-like

Score

Hard background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH

EVALUATE



0ms

1.5ms

4ms

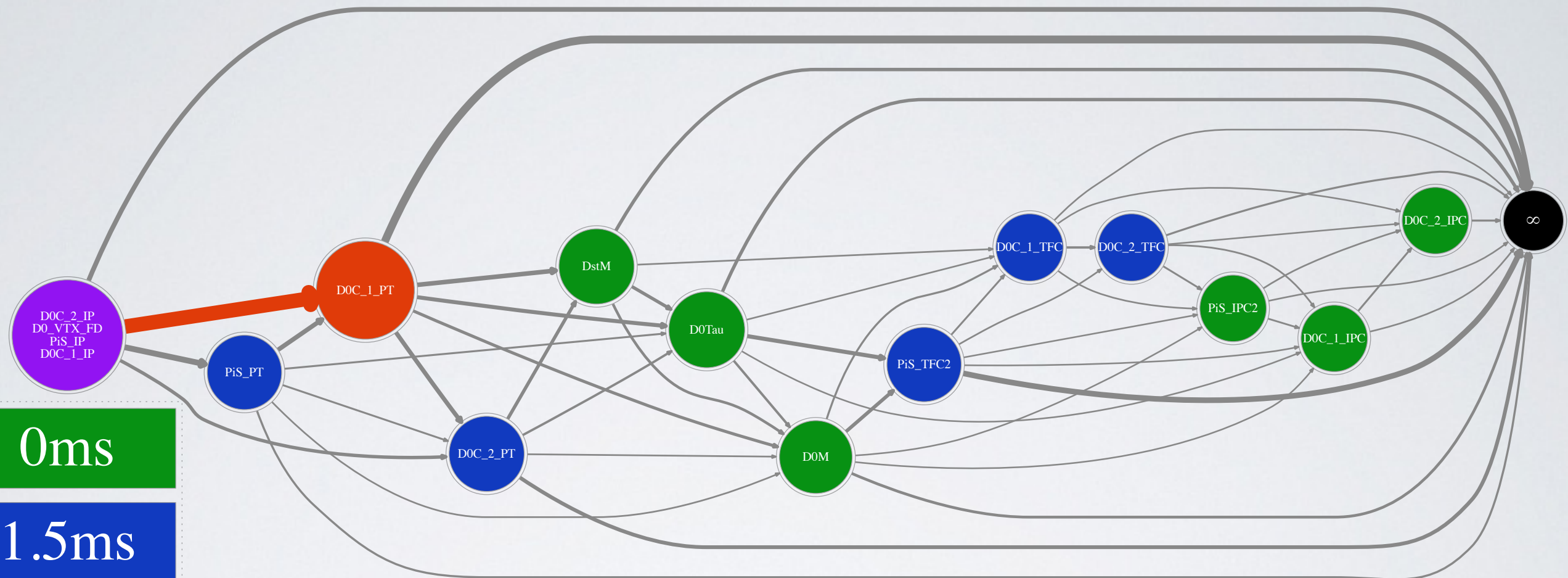
Background-like

Signal-like

Score

Hard background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



0ms

1.5ms

4ms

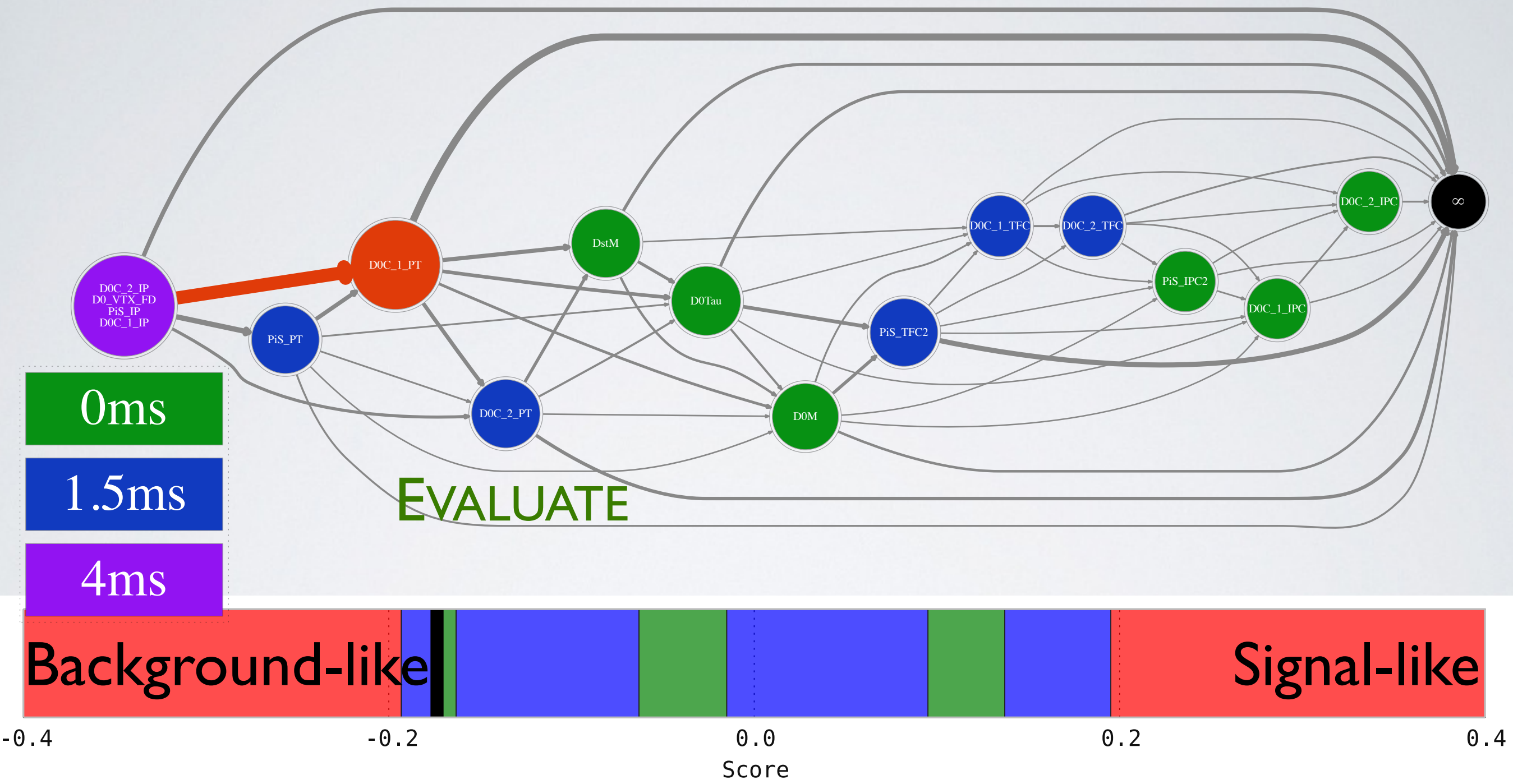
Background-like

Signal-like

Score

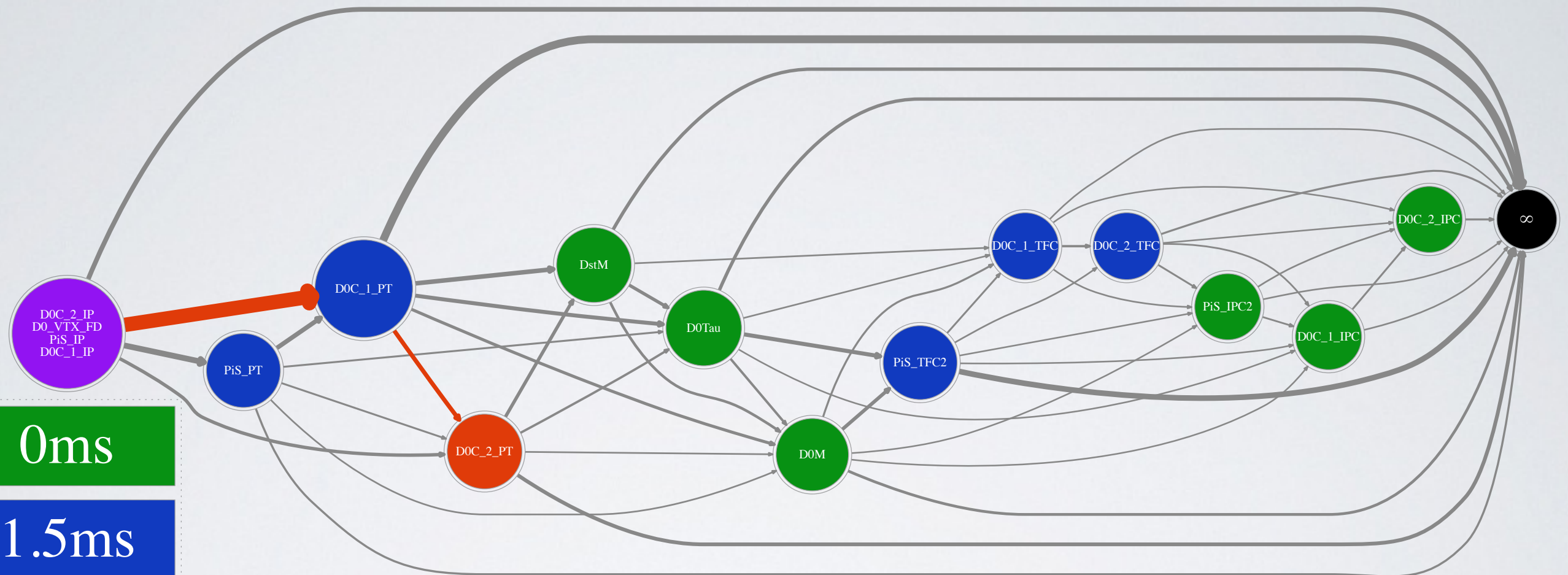
Hard background

MDDAG: A SIGNAL/BCKKG DECISION GRAPH



Hard background

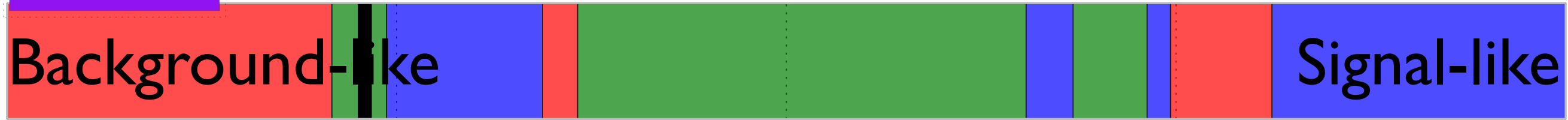
MDDAG: A SIGNAL/BCKGG DECISION GRAPH



0ms

1.5ms

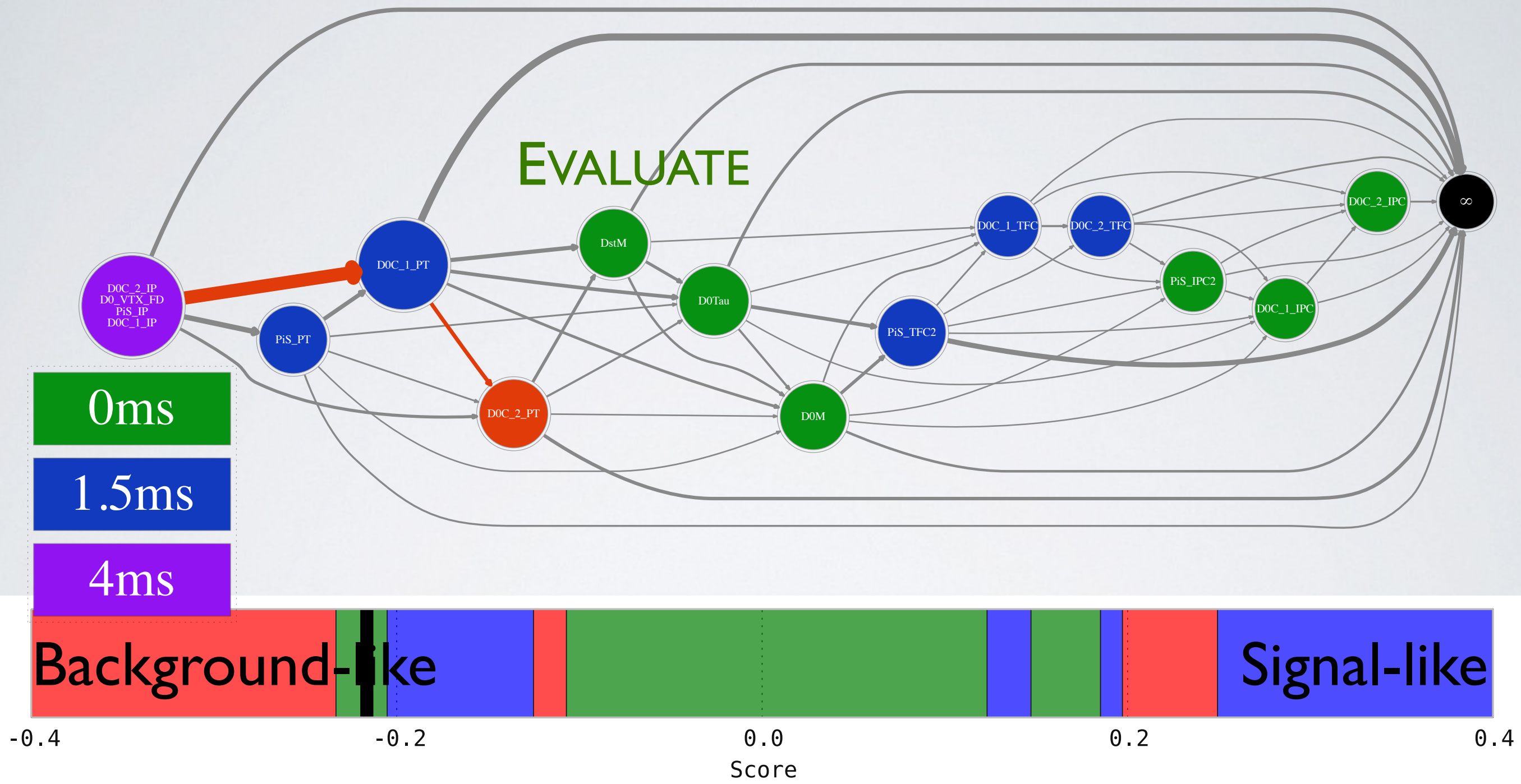
4ms



-0.4 -0.2 0.0 0.2 0.4
Score

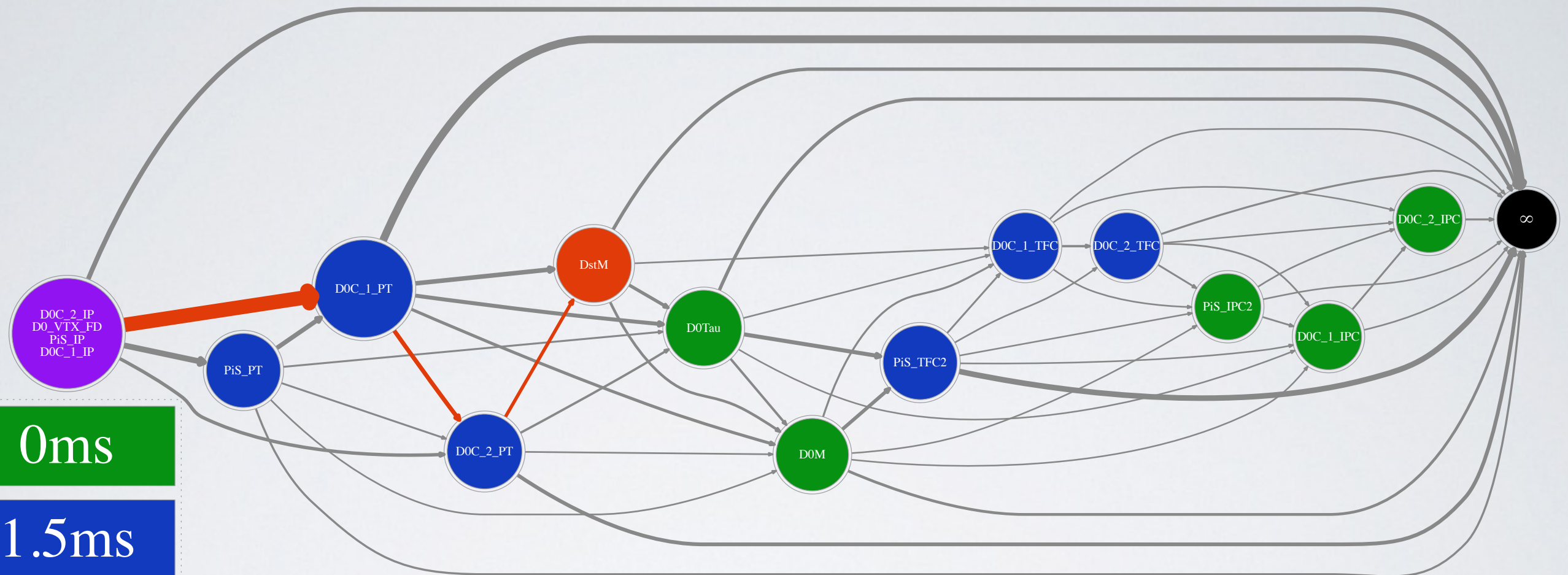
Hard background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



Hard background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



0ms

1.5ms

4ms

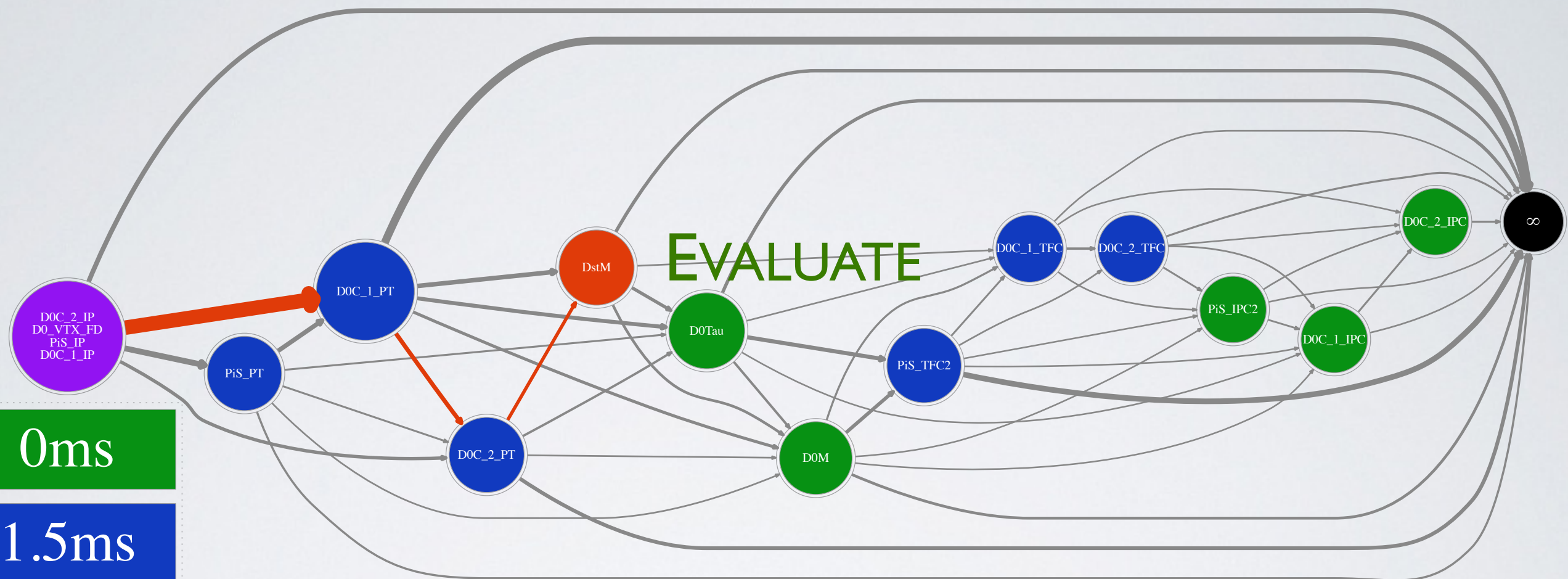
Background-like

Signal-like

Score

Hard background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



0ms

1.5ms

4ms

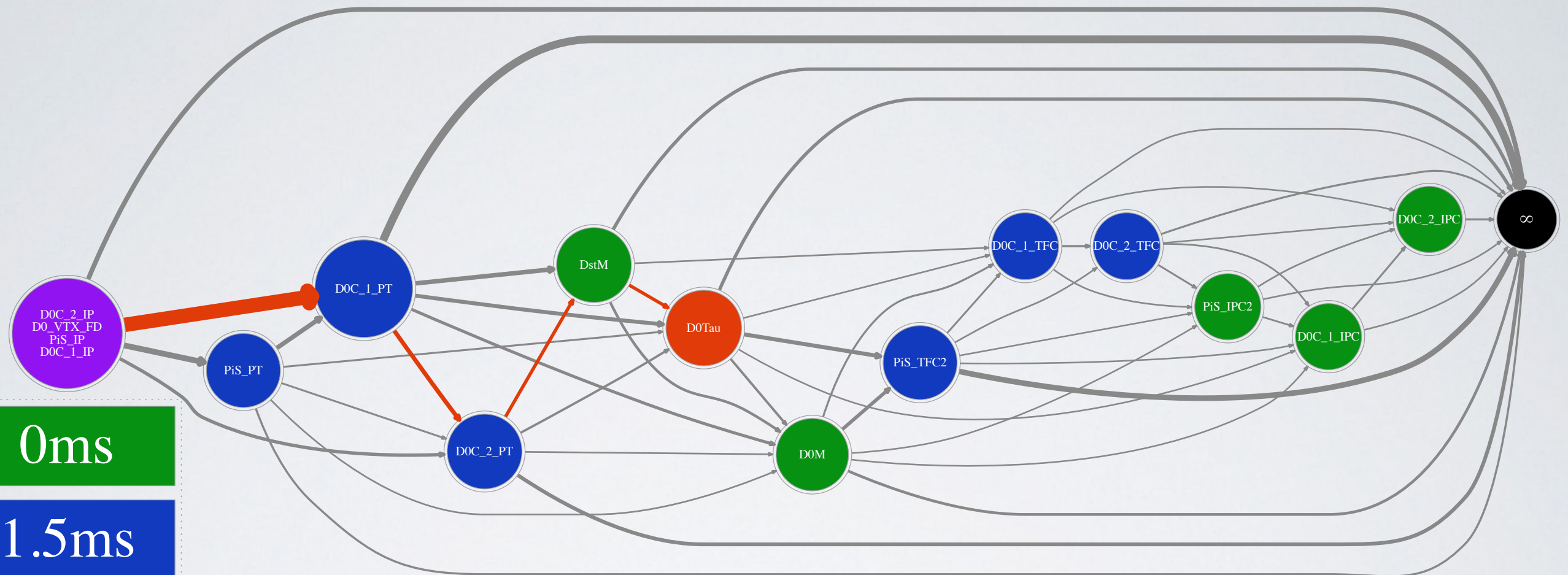
Background-like

Signal-like

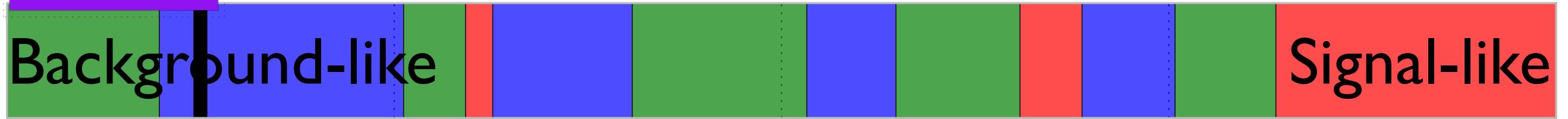
Score

Hard background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



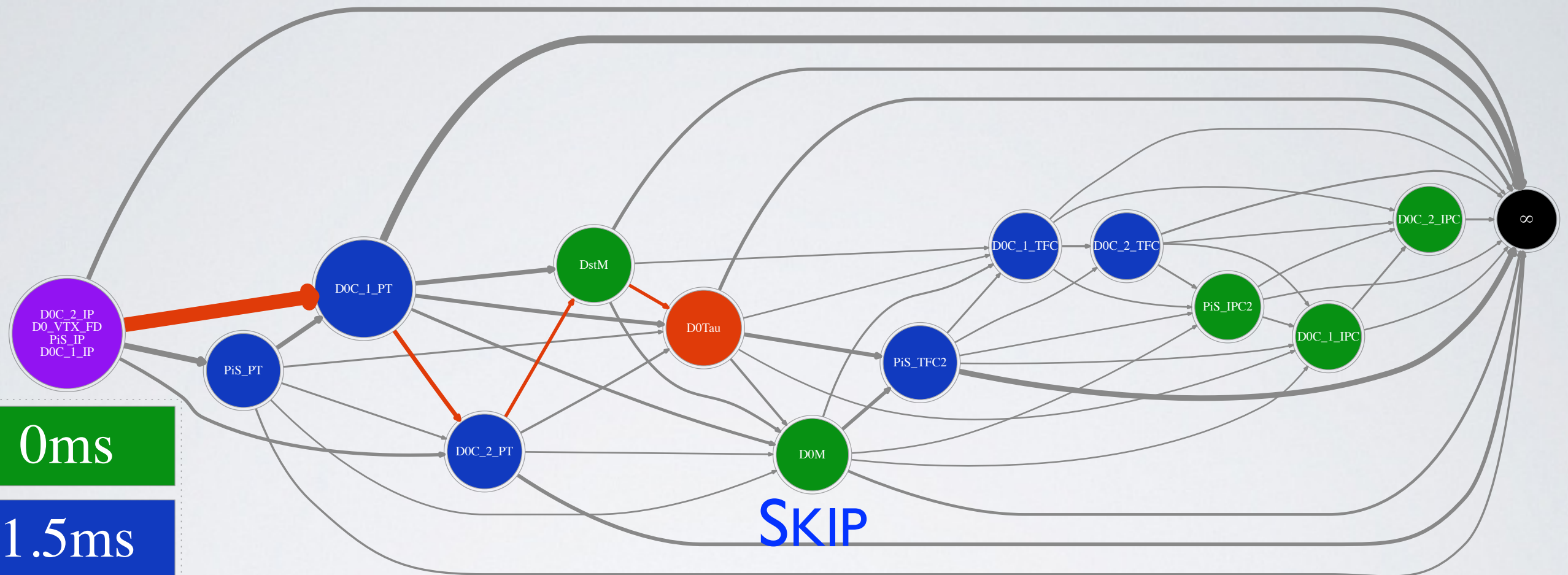
0ms
1.5ms
4ms



Score

Hard background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



0ms

1.5ms

4ms

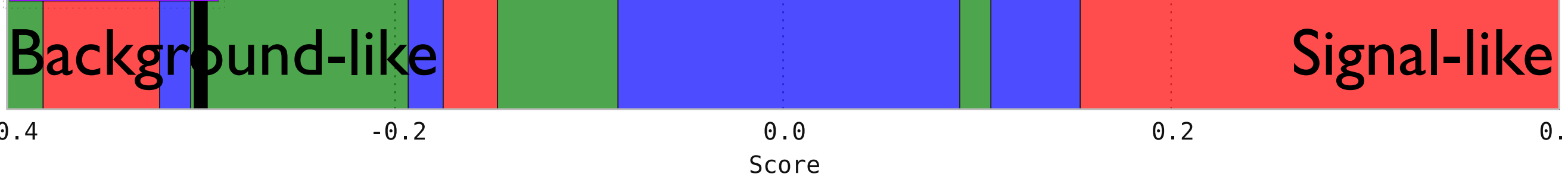
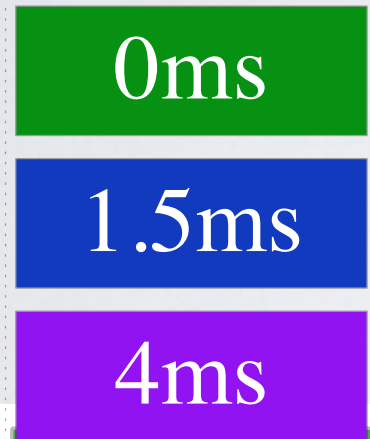
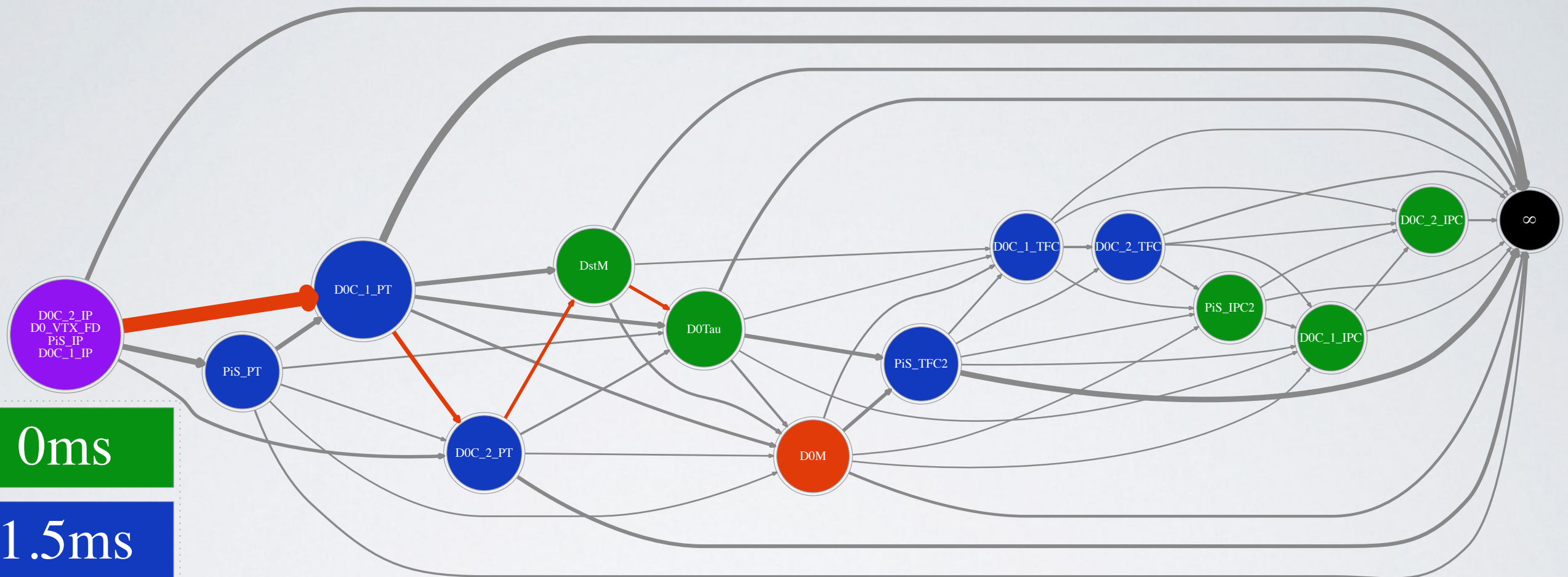
Background-like

Signal-like

Score

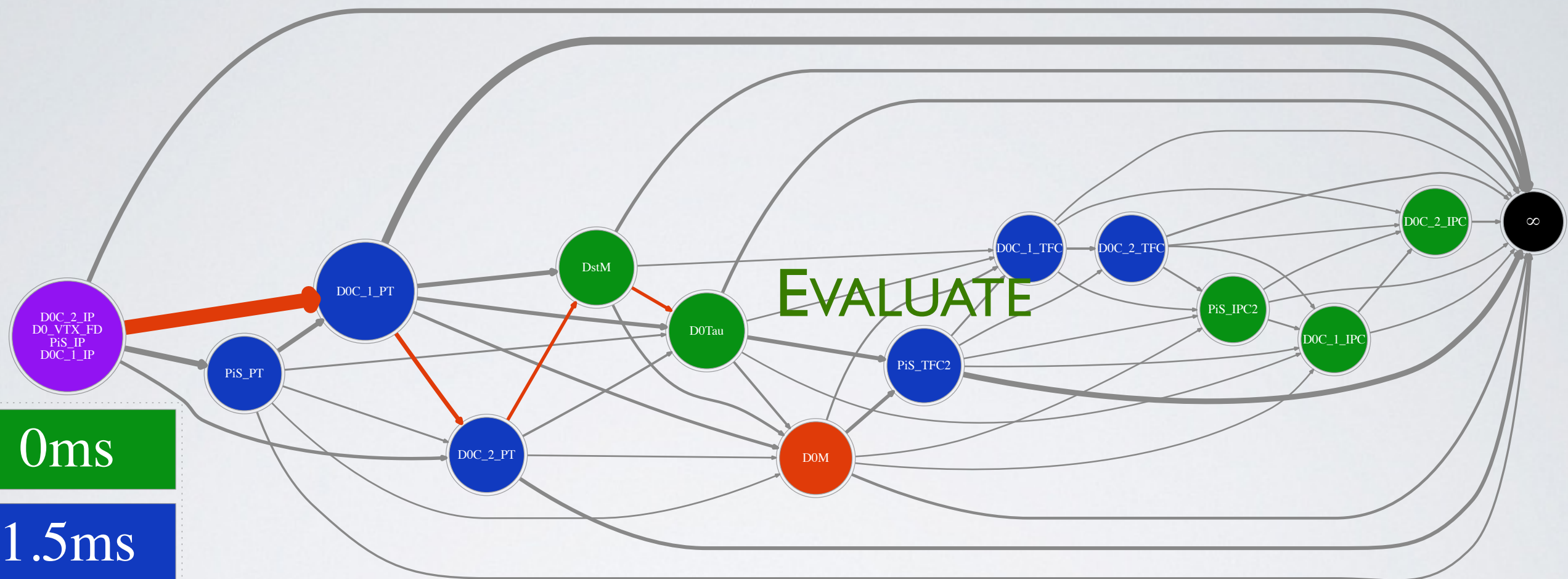
Hard background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH

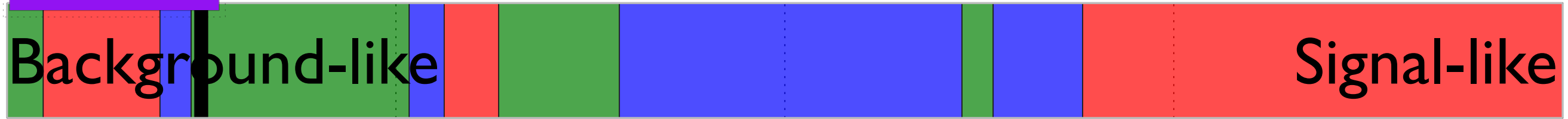


Hard background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



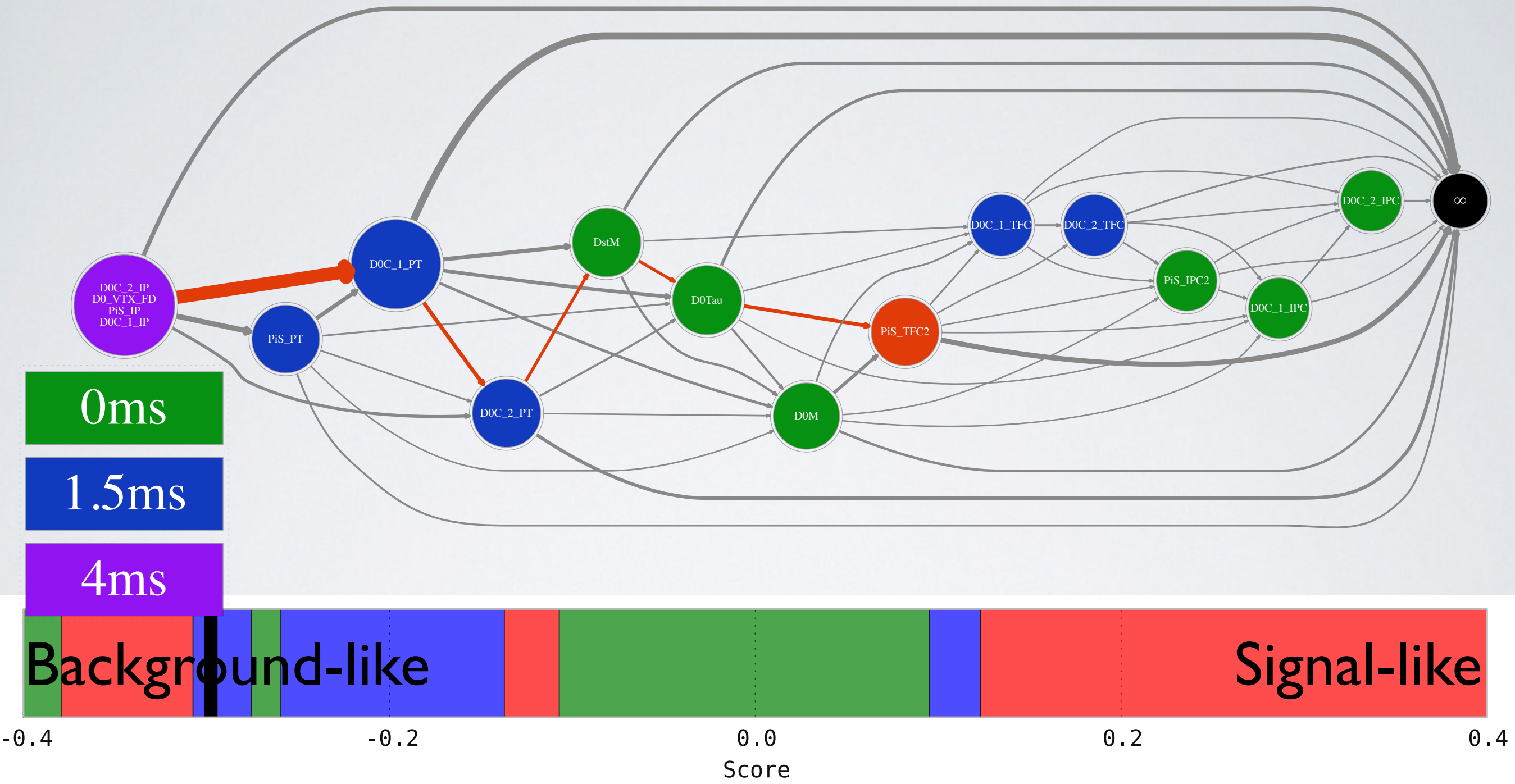
0ms
1.5ms
4ms



-0.4 -0.2 0.0 0.2 0.4
Score

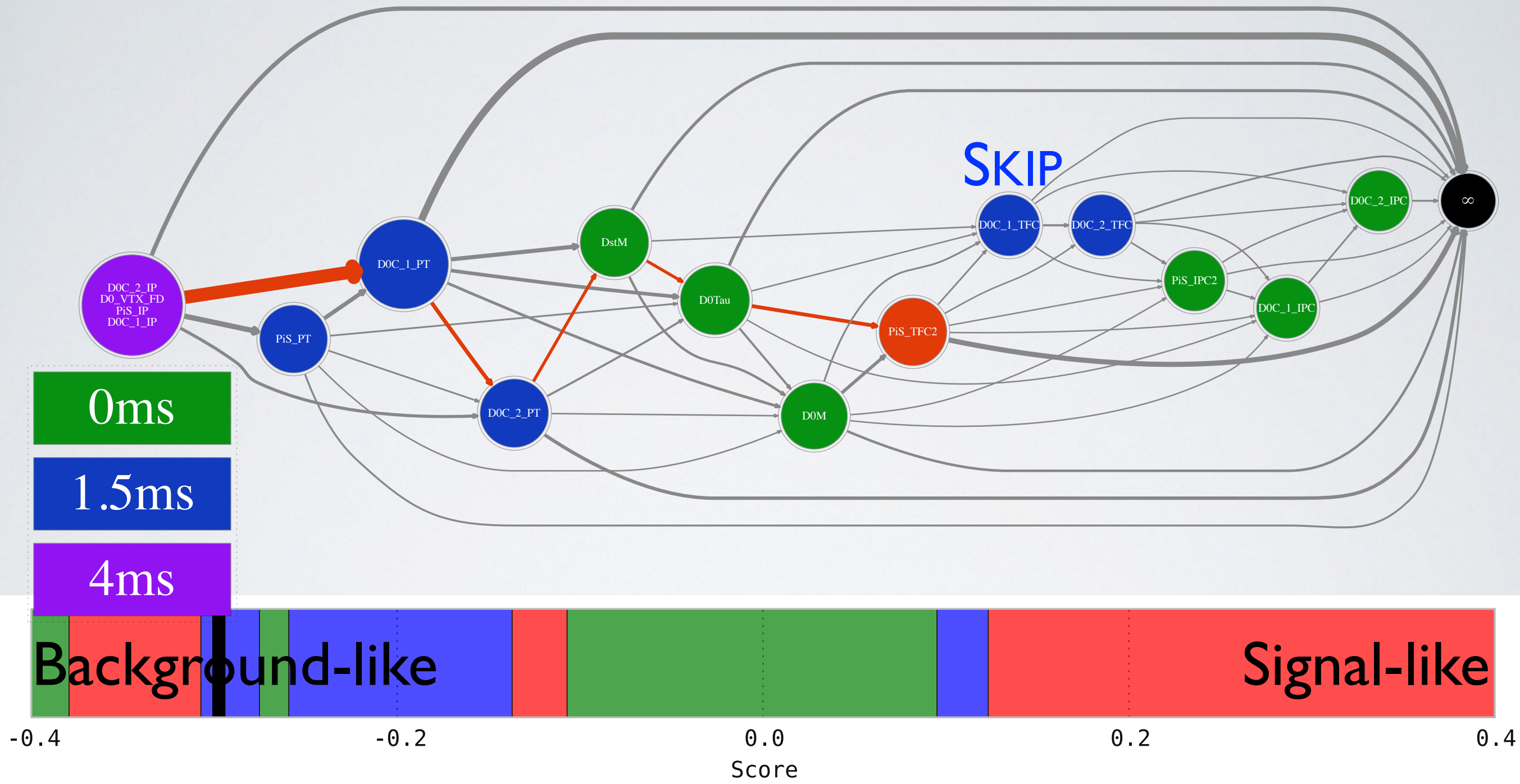
Hard background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



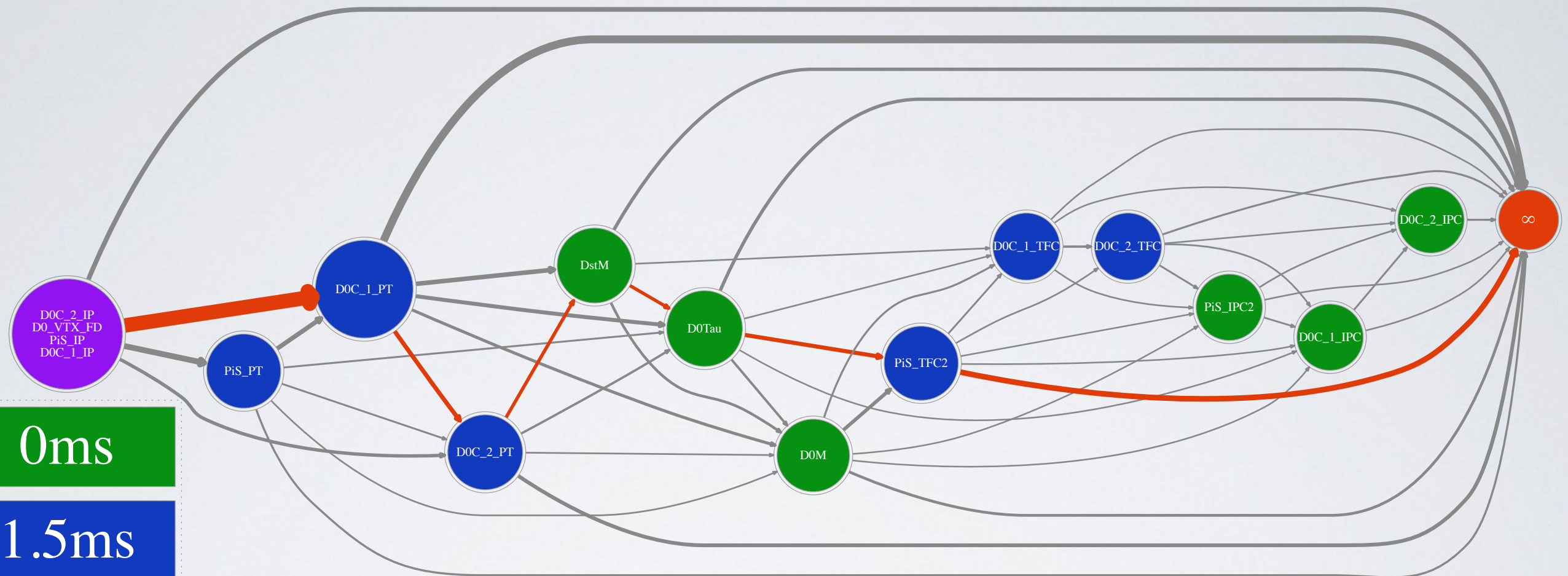
Hard background

MDDAG: A SIGNAL/BCKKG DECISION GRAPH



Hard background

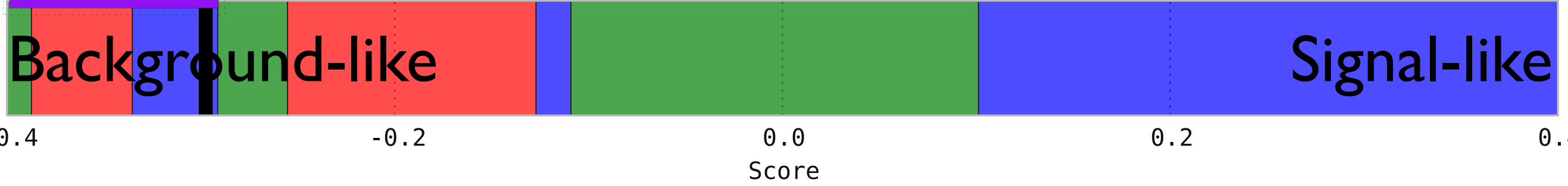
MDDAG: A SIGNAL/BCKKG DECISION GRAPH



0ms

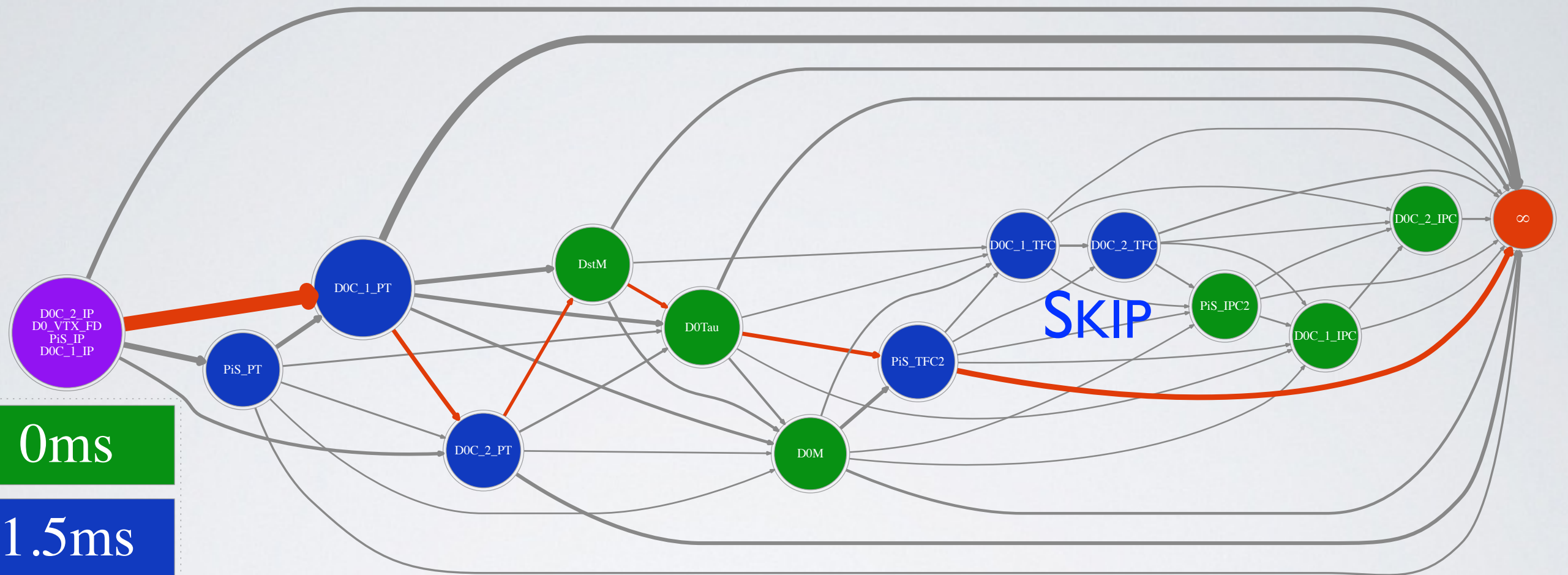
1.5ms

4ms



Hard background

MDDAG: A SIGNAL/BCKG DECISION GRAPH



0ms

1.5ms

4ms

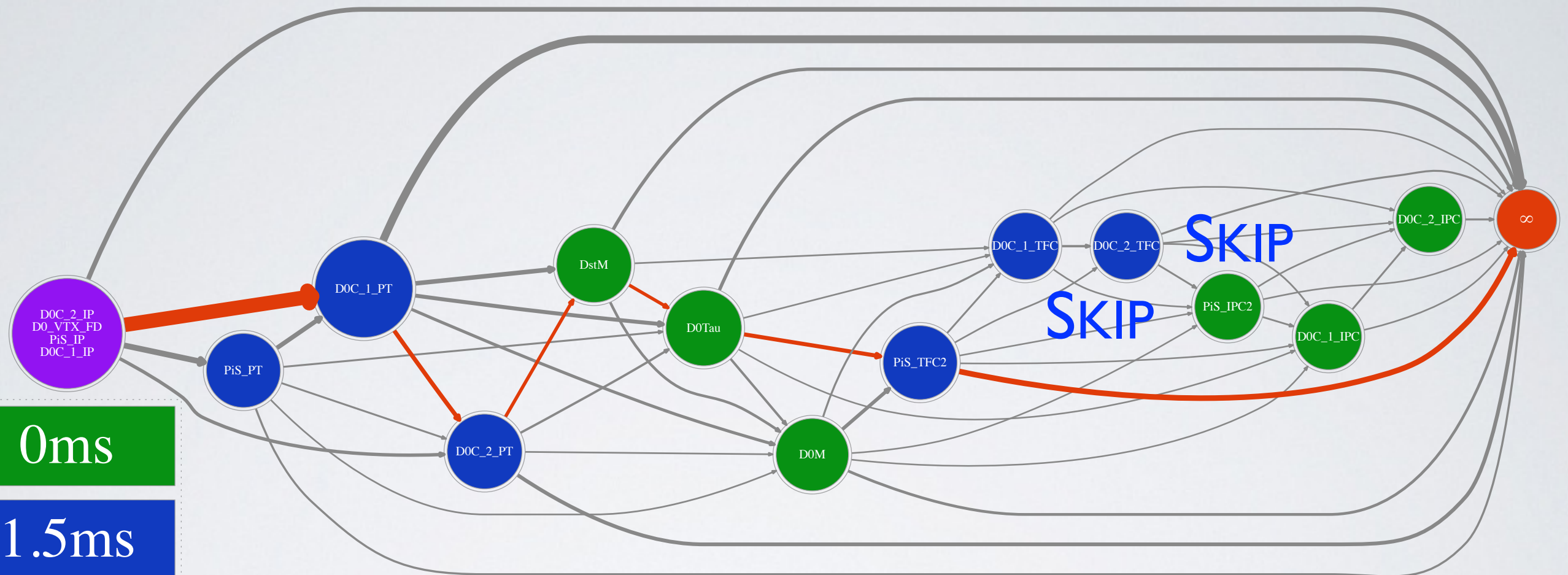
Background-like

Signal-like

Score

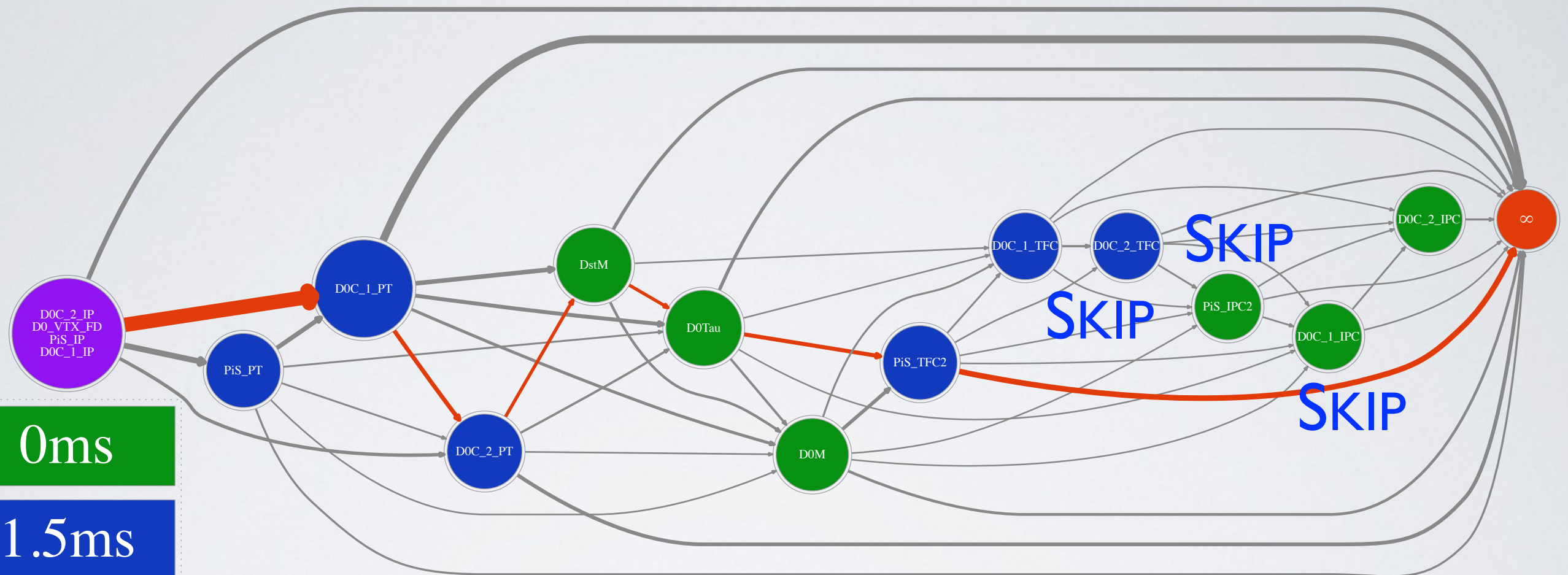
Hard background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



Hard background

MDDAG: A SIGNAL/BCKG DECISION GRAPH



0ms

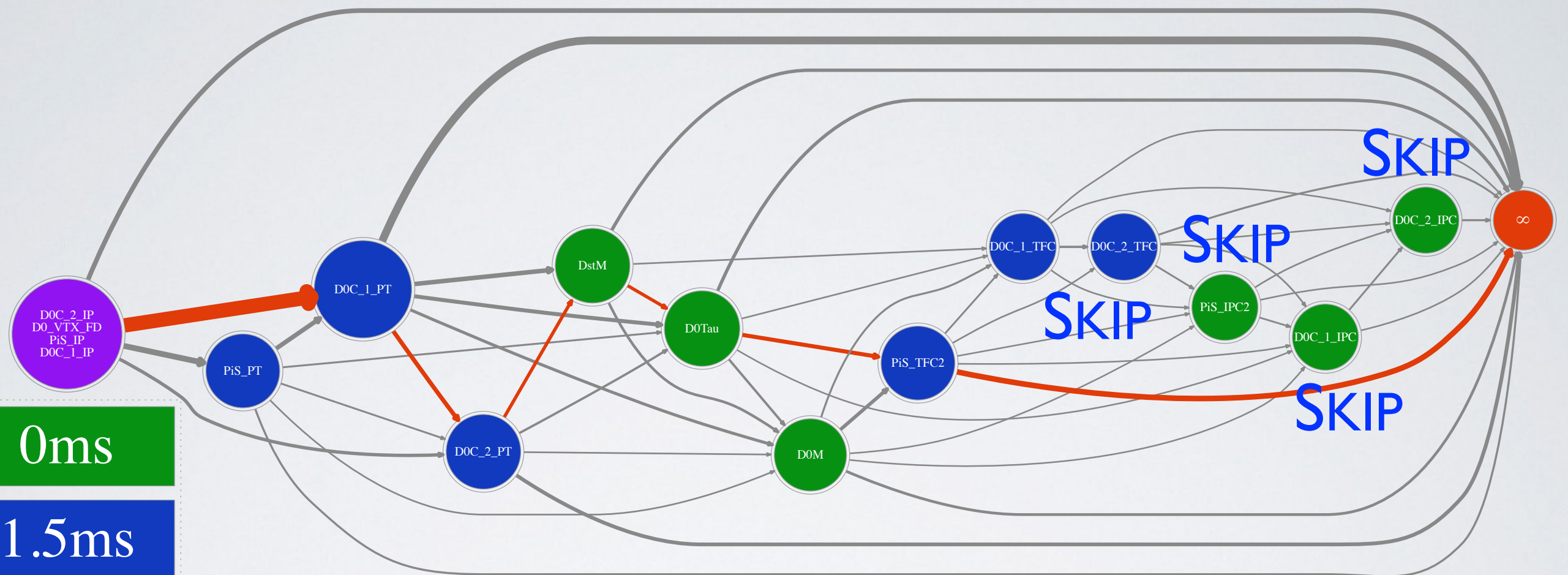
1.5ms

4ms

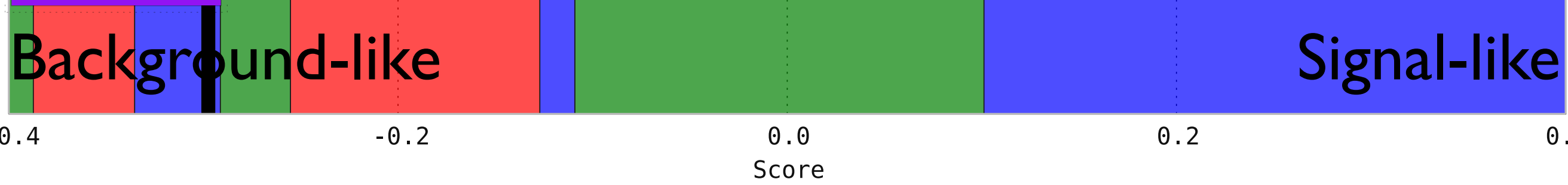


Hard background

MDDAG: A SIGNAL/BCKG DECISION GRAPH

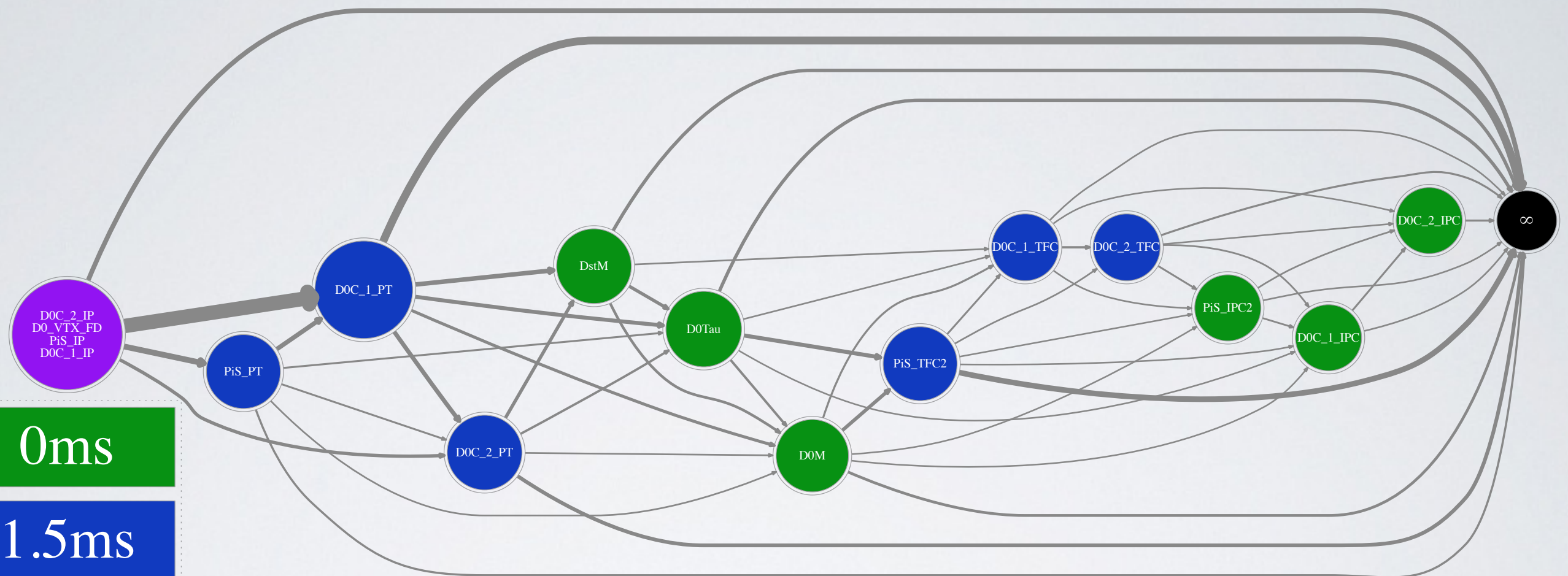


0ms
1.5ms
4ms



Hard background

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



0ms

1.5ms

4ms

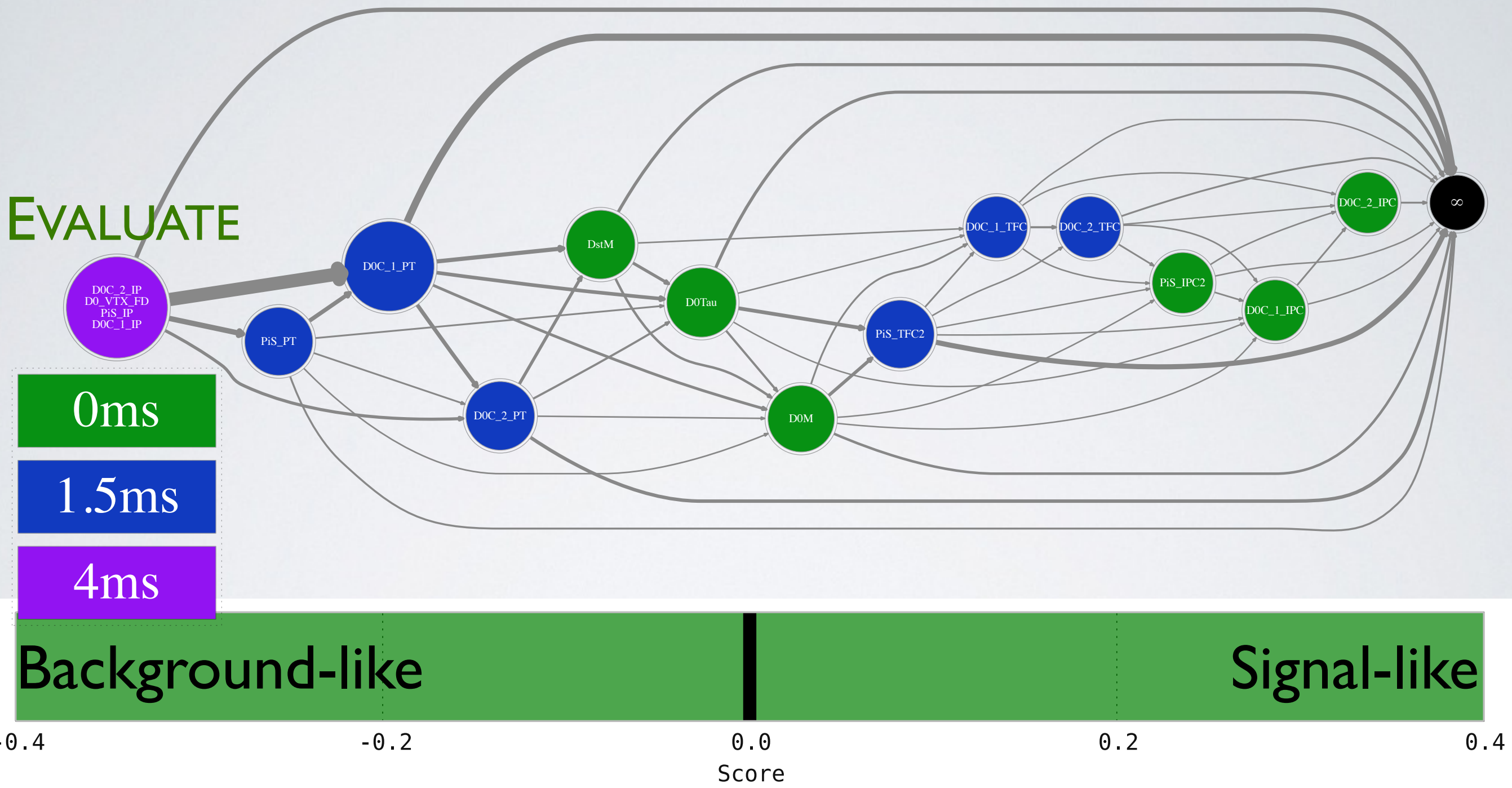
Background-like

Signal-like

Score

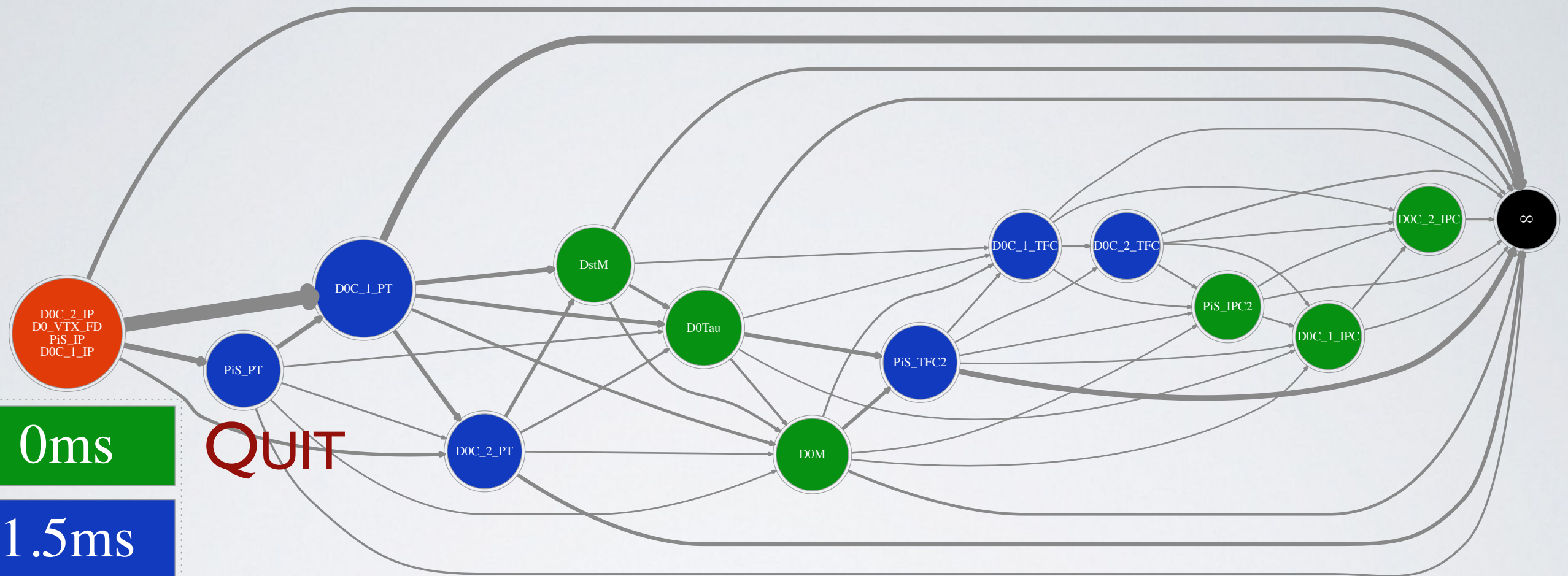
Easy signal

MDDAG: A SIGNAL/BCKG DECISION GRAPH



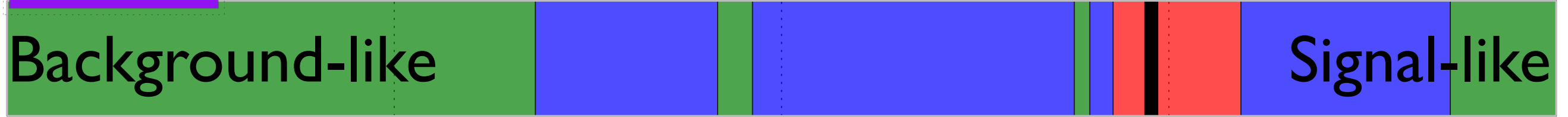
Easy signal
36

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



0ms
1.5ms
4ms

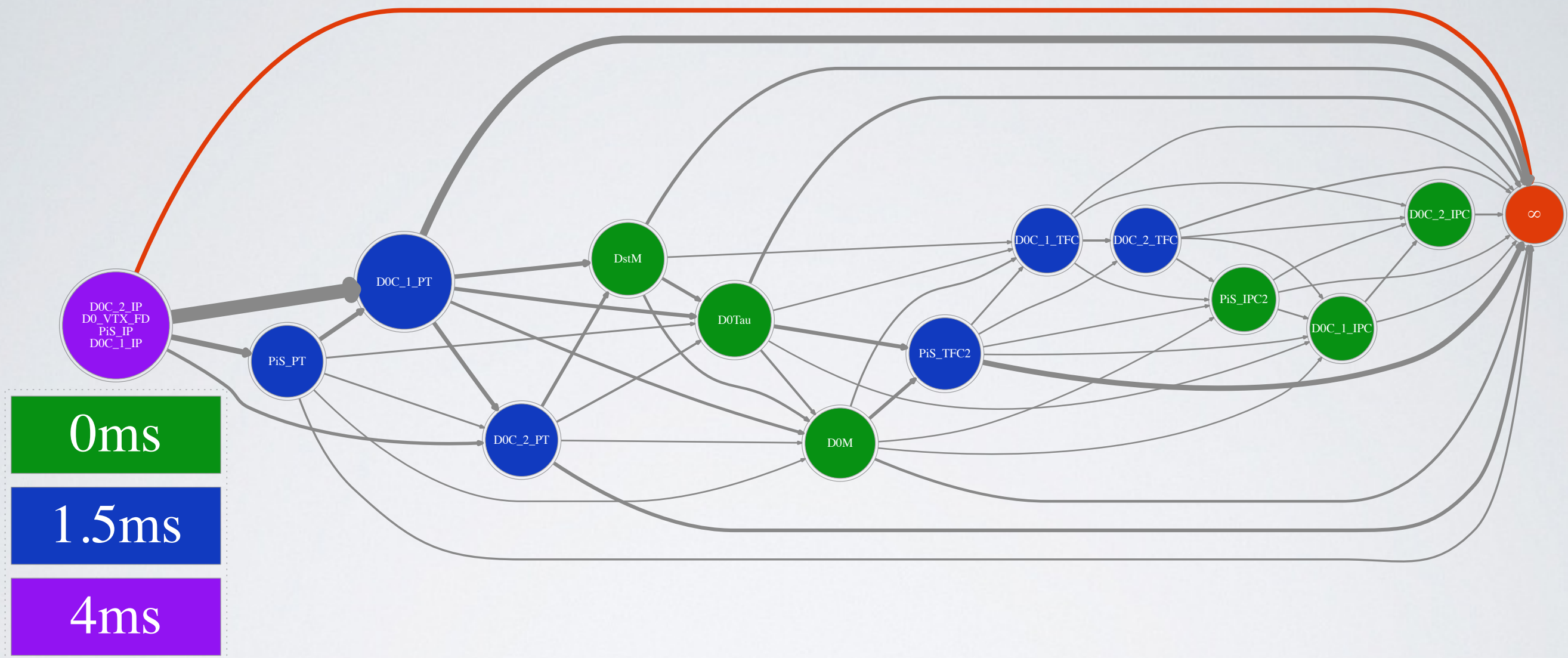
QUIT



-0.4 -0.2 0.0 0.2 0.4
Score

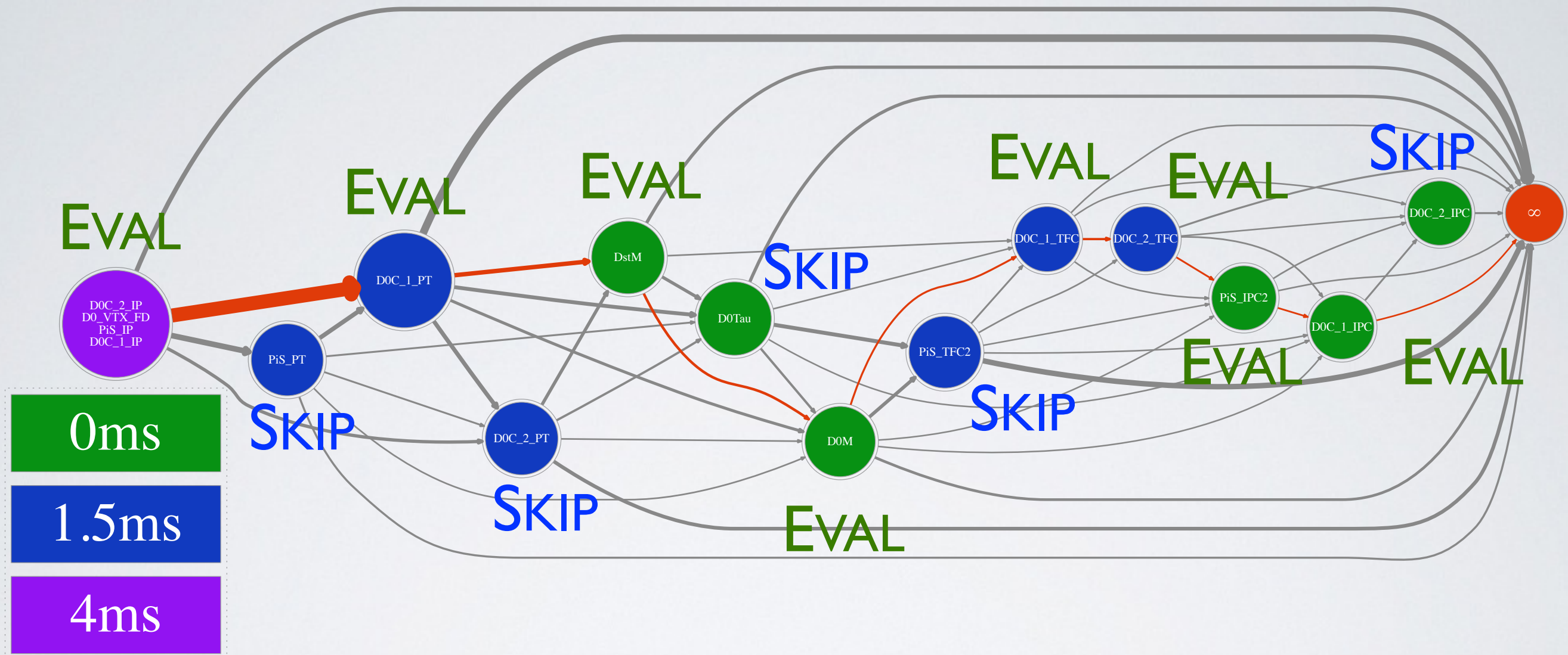
Easy signal

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



Easy signal₃₈

MDDAG: A SIGNAL/BCKGG DECISION GRAPH



Hard signal

BUDGETED CLASSIFICATION

- Classification with **test-time constraints**
- An **active research** area due to IT applications
- To be exploited for **trigger design**

BUDGETED CLASSIFICATION

Take-home message:

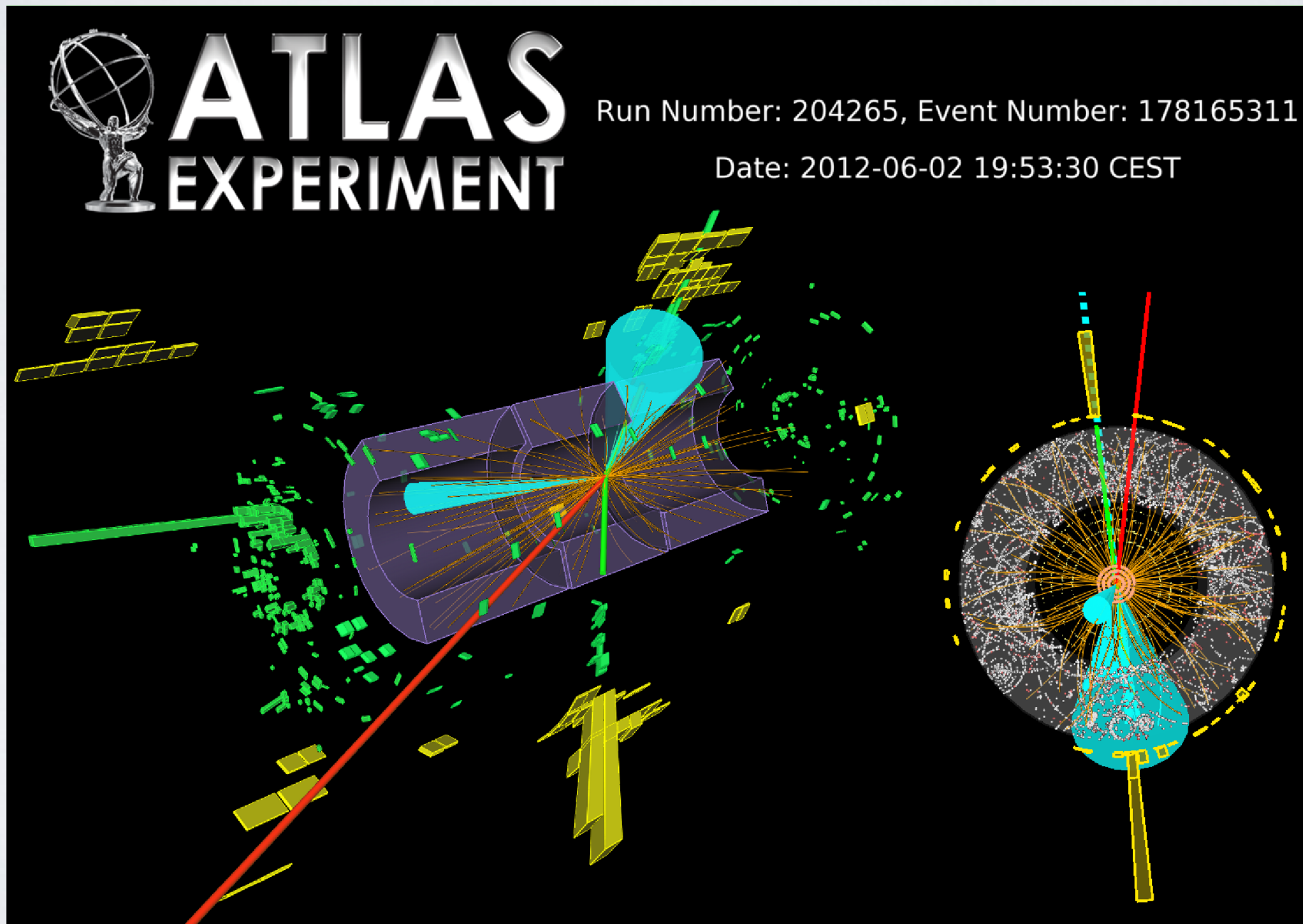
Once you go software trigger,
the set of **possible solutions opens up**

HEP-INSPIRED OBJECTIVES

- Low **error probability** ~ high **accuracy** ~ high **efficiency**
- **Fast** classification
- More exotic goals (or constraints):
 - **easy control** of detection **efficiency** in different signal or background classes
 - **efficiency unbiased** in certain variables
 - feeding the classifier into a statistical (counting) test: **maximize test sensitivity**
 - in the big data regime **systematics** are becoming more important than statistical efficiency

CLASSIFICATION FOR DISCOVERY

The HiggsML challenge



CLASSIFICATION FOR DISCOVERY

CLASSIFICATION FOR DISCOVERY

- In a nutshell

CLASSIFICATION FOR DISCOVERY

- In a nutshell
 - A vector x of variables is extracted from each event

CLASSIFICATION FOR DISCOVERY

- In a nutshell
 - A vector \mathbf{x} of variables is extracted from each event
 - A classifier $g(\mathbf{x})$ is trained to separate signal from background

CLASSIFICATION FOR DISCOVERY

- In a nutshell
 - A vector \mathbf{x} of variables is extracted from each event
 - A classifier $g(\mathbf{x})$ is trained to separate signal from background
 - The background b is estimated in the selection region
 $G = \{\mathbf{x} : g(\mathbf{x}) = s\}$

CLASSIFICATION FOR DISCOVERY

- In a nutshell
 - A vector \mathbf{x} of variables is extracted from each event
 - A classifier $g(\mathbf{x})$ is trained to separate signal from background
 - The background b is estimated in the selection region
 $G = \{\mathbf{x} : g(\mathbf{x}) = s\}$
 - Discovery is made when the number of real events n is significantly higher than b

CLASSIFICATION FOR DISCOVERY

CLASSIFICATION FOR DISCOVERY

- Exciting **physics**

CLASSIFICATION FOR DISCOVERY

- Exciting **physics**
 - The **Higgs to tau-tau** excess is **not yet at five sigma**
Tech. Rep. ATLAS-CONF-2013-108

CLASSIFICATION FOR DISCOVERY

- Exciting **physics**
 - The **Higgs to tau-tau** excess is **not yet at five sigma**
Tech. Rep. ATLAS-CONF-2013-108
- Exciting **data science** (statistics and machine learning)

CLASSIFICATION FOR DISCOVERY

- Exciting **physics**
 - The **Higgs to tau-tau** excess is **not yet at five sigma**
Tech. Rep. ATLAS-CONF-2013-108
- Exciting **data science** (statistics and machine learning)
 - What is the **theoretical relationship** between **classification** and **test sensitivity**?

CLASSIFICATION FOR DISCOVERY

- Exciting **physics**
 - The **Higgs to tau-tau** excess is **not yet at five sigma**
Tech. Rep. ATLAS-CONF-2013-108
- Exciting **data science** (statistics and machine learning)
 - What is the **theoretical relationship** between **classification** and **test sensitivity**?
 - What is the **quantitative criteria** to optimize?

CLASSIFICATION FOR DISCOVERY

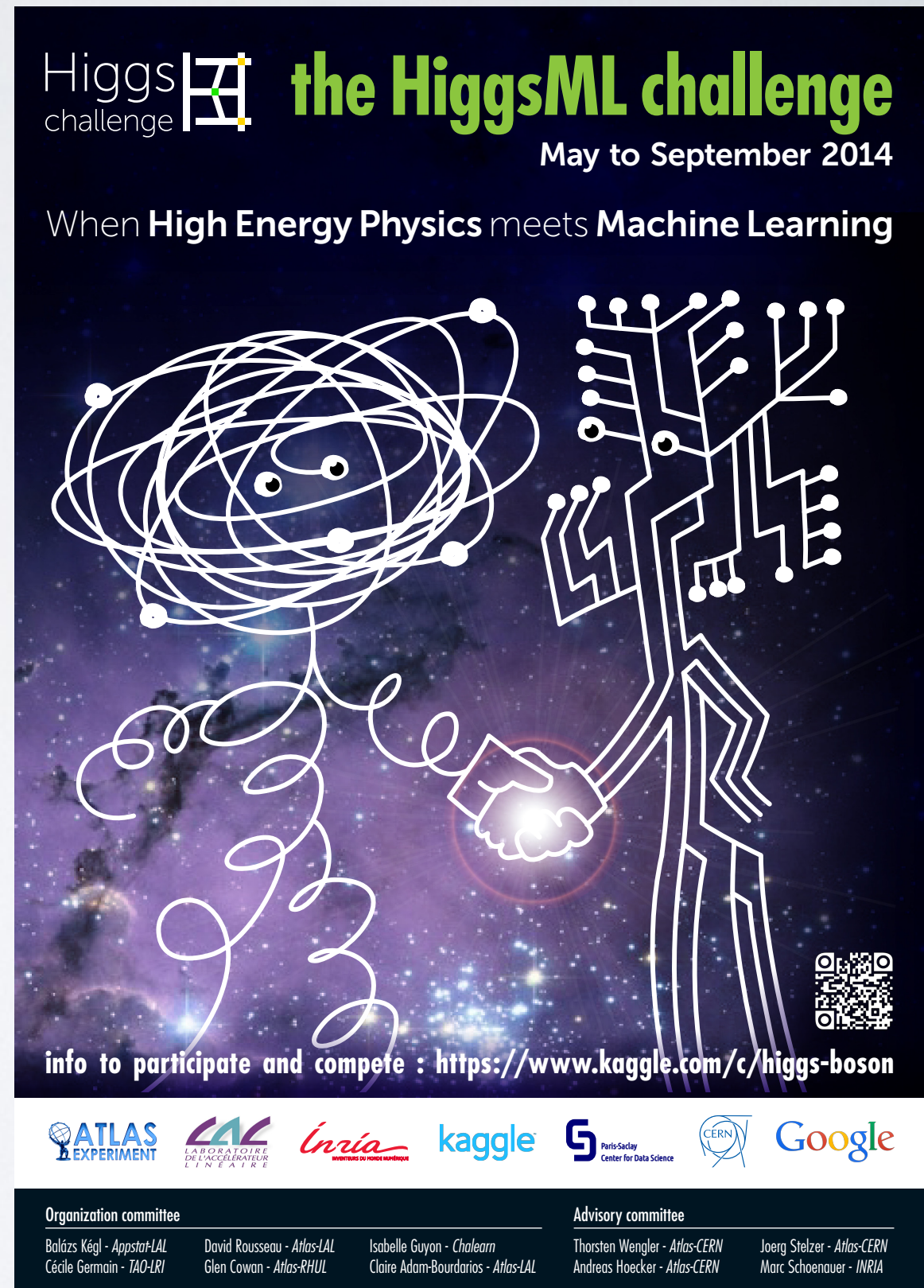
- Exciting **physics**
 - The **Higgs to tau-tau** excess is **not yet at five sigma**
Tech. Rep. ATLAS-CONF-2013-108
- Exciting **data science** (statistics and machine learning)
 - What is the **theoretical relationship** between **classification** and **test sensitivity**?
 - What is the **quantitative criteria** to optimize?
 - How to formally include **systematic uncertainties**?

CLASSIFICATION FOR DISCOVERY


- Exciting **physics**
 - The **Higgs to tau-tau** excess is **not yet at five sigma**
Tech. Rep. ATLAS-CONF-2013-108
- Exciting **data science** (statistics and machine learning)
 - What is the **theoretical relationship** between **classification** and **test sensitivity**?
 - What is the **quantitative criteria** to optimize?
 - How to formally include **systematic uncertainties**?
 - Can we **redesign classical algorithms** (boosting, SVM, neural nets) for optimizing this criteria?

CLASSIFICATION FOR DISCOVERY








We are running a
data challenge
to answer some of these
questions



The poster for the HiggsML challenge features a central illustration of two figures shaking hands against a starry space background. The figure on the left is composed of white lines forming an atomic model with a central nucleus and orbiting electrons. The figure on the right is a stylized tree-like structure made of white circuit board traces. The text at the top reads 'Higgs challenge' with a logo, followed by 'the HiggsML challenge' in green, and 'May to September 2014'. Below this is the tagline 'When High Energy Physics meets Machine Learning'. At the bottom, there is a QR code and the URL 'https://www.kaggle.com/c/higgs-boson'. The footer contains logos for ATLAS EXPERIMENT, LAL (LABORATOIRE DE L'ACCELERATEUR LINEAIRE), Inria, Kaggle, Paris-Saclay Center for Data Science, CERN, and Google.

Higgs challenge  **the HiggsML challenge**
May to September 2014
When **High Energy Physics** meets **Machine Learning**

info to participate and compete : <https://www.kaggle.com/c/higgs-boson>

Organization committee
Balázs Kégl - *Appsta-LAL* David Rousseau - *Atlas-LAL* Isabelle Guyon - *Chaleam*
Cécile Germain - *TAO-LRI* Glen Cowan - *Atlas-RHUL* Claire Adam-Bourdarios - *Atlas-LAL*

Advisory committee
Thorsten Wengler - *Atlas-CERN* Joerg Stelzer - *Atlas-CERN*
Andreas Hoecker - *Atlas-CERN* Marc Schoenauer - *INRIA*

CLASSIFICATION FOR DISCOVERY

The formal setup

- We **simulate** data: $\mathcal{D} = \{(\mathbf{x}_1, y_1, w_1), \dots, (\mathbf{x}_n, y_n, w_n)\}$
 - $\mathbf{x}_i \in \mathbb{R}^d$ is the **feature vector**
 - $y_i \in \{\mathbf{b}_{\text{background}}, \mathbf{s}_{\text{Signal}}\}$ is the **label**
 - $w_i \in \mathbb{R}^+$ is a non-negative **weight** (importance sampling)
 - let $\mathcal{S} = \{i : y_i = \mathbf{s}\}$ and $\mathcal{B} = \{i : y_i = \mathbf{b}\}$ be the **index sets** of signal and background events, respectively
- Maximize the **Approximate Median Significance**


G. Cowan, K. Cranmer, E. Gross, and O. Vitells. EPJ C, 71:1554, 2011.

$$\text{AMS} = \sqrt{2 \left((s + b) \ln \left(1 + \frac{s}{b} \right) - s \right)} \approx \frac{s}{\sqrt{b}}$$

- $\hat{\mathcal{G}} = \{i : g(\mathbf{x}_i) = \mathbf{s}\}$
- $s = \sum_{i \in \mathcal{S} \cap \hat{\mathcal{G}}} w_i$
- $b = \sum_{i \in \mathcal{B} \cap \hat{\mathcal{G}}} w_i$

CLASSIFICATION FOR DISCOVERY


A tool for getting
the **ML community** excited
about your problem




the HiggsML challenge

May to September 2014

When **High Energy Physics** meets **Machine Learning**



info to participate and compete : <https://www.kaggle.com/c/higgs-boson>




ATLAS EXPERIMENT **LAL LABORATOIRE DE L'ACCELERATEUR LINEAIRE** **Inria INSTITUTS DU FORDA MATHÉMATIQUE** **kaggle** **Paris-Saclay Center for Data Science** **CERN** **Google**

Organization committee
Balázs Kégl - *Appsta-LAL* David Rousseau - *Atlas-LAL* Isabelle Guyon - *Chaleam*
Cécile Germain - *TAO-LRI* Glen Cowan - *Atlas-RHUL* Claire Adam-Bourdarios - *Atlas-LAL*

Advisory committee
Thorsten Wengler - *Atlas-CERN* Joerg Stelzer - *Atlas-CERN*
Andreas Hoecker - *Atlas-CERN* Marc Schoenauer - *INRIA*

CLASSIFICATION FOR DISCOVERY


- Official **ATLAS GEANT4** simulations
 - **30 features** (variables)
 - **250K training**: input, label, weight
 - **100K public test** (AMS displayed real-time), only input
 - **450K private test** (to determine the winner after the closing of the challenge), only input
 - public and private tests set are **shuffled**, participants submit a vector of **550K** labels




the HiggsML challenge

May to September 2014

When **High Energy Physics** meets **Machine Learning**



info to participate and compete : <https://www.kaggle.com/c/higgs-boson>



ATLAS EXPERIMENT **LAL** LABORATOIRE DE L'ACCELERATEUR LINEAIRE **Inria** INSTITUTS DU FORDA NUMERIQUE **kaggle** Paris-Saclay Center for Data Science **CERN** **Google**


Organization committee
Balázs Kégl - *Appsta-LAL* David Rousseau - *Atlas-LAL* Isabelle Guyon - *Chaleam*
Cécile Germain - *TAO-LRI* Glen Cowan - *Atlas-RHUL* Claire Adam-Bourdarios - *Atlas-LAL*

Advisory committee
Thorsten Wengler - *Atlas-CERN* Joerg Stelzer - *Atlas-CERN*
Andreas Hoecker - *Atlas-CERN* Marc Schoenauer - *INRIA*

CLASSIFICATION FOR DISCOVERY

- **16K\$** prize pool
 - **7-4-2K\$** for the **three top participants**
 - **HEP meets ML award** for the most useful model, decided by the ATLAS members of the organizing committee



Higgs challenge  **the HiggsML challenge**
May to September 2014
When **High Energy Physics** meets **Machine Learning**

info to participate and compete : <https://www.kaggle.com/c/higgs-boson>

Organization committee: Balázs Kégl - *Appsta-LAL*, Cécile Germain - *TAO-LRI*, David Rousseau - *Atlas-LAL*, Glen Cowan - *Atlas-RHUL*, Isabelle Guyon - *Chaleam*, Claire Adam-Bourdarios - *Atlas-LAL*

Advisory committee: Thorsten Wengler - *Atlas-CERN*, Andreas Hoecker - *Atlas-CERN*, Joerg Stelzer - *Atlas-CERN*, Marc Schoenauer - *INRIA*

CLASSIFICATION FOR DISCOVERY

kaggle

Customer Solutions

Competitions

Community ▾

Sign up

Login



\$13,000 • 1,627 teams

Higgs Boson Machine Learning Challenge

Mon 12 May 2014

Mon 15 Sep 2014 (11 days to go)

Enter/Merge by

Dashboard ▾

Leaderboard - Higgs Boson Machine Learning Challenge

This leaderboard is calculated on approximately 18% of the test data. The final results will be based on the other 82%, so the final standings may be different.

See someone using multiple accounts? [Let us know.](#)

#	Δ1w	Team Name <small>‡ model uploaded * in the money</small>	Score <small>?</small>	Entries	Last Submission UTC (Best - Last Submission)
1	—	Gábor Melis *	3.85059	97	Thu, 04 Sep 2014 08:07:20 (-34d)
2	↑1	Luboš Motl's team <small>👤</small> *	3.84522	535	Thu, 04 Sep 2014 09:22:47 (-1.9h)
3	↓1	Tim Salimans *	3.84428	46	Tue, 12 Aug 2014 17:42:01 (-12.2d)

CLASSIFICATION FOR DISCOVERY

kaggle

Customer Solutions

Competitions

Community ▾

Sign up

Login



\$13,000 **1,627 teams**

Higgs Boson Machine Learning Challenge

Mon 12 May 2014

Mon 15 Sep 2014 (11 days to go)

Enter/Merge by

Dashboard ▾


Leaderboard - Higgs Boson Machine Learning Challenge

This leaderboard is calculated on approximately 18% of the test data. The final results will be based on the other 82%, so the final standings may be different.


See someone using multiple accounts? [Let us know.](#)

#	Δ1w	Team Name <small>‡ model uploaded * in the money</small>	Score <small>?</small>	Entries	Last Submission UTC (Best - Last Submission)
1	—	Gábor Melis *	3.85059	97	Thu, 04 Sep 2014 08:07:20 (-34d)
2	↑1	Luboš Motl's team <small>👤</small> *	3.84522	535	Thu, 04 Sep 2014 09:22:47 (-1.9h)
3	↓1	Tim Salimans *	3.84428	46	Tue, 12 Aug 2014 17:42:01 (-12.2d)


CLASSIFICATION FOR DISCOVERY

#	Δ1w	Team Name <small>‡ model uploaded * in the MONEY</small>	Score	Entries	Last Submission UTC (Best - Last Submission)
1	—	Gábor Melis *	3.85059	97	Thu, 04 Sep 2014 08:07:20 (-34d)
2	↑1	Luboš Motl's team *	3.84522	535	Thu, 04 Sep 2014 09:22:47 (-1.9h)
3	↓1	Tim Salimans *	3.84428	46	Tue, 12 Aug 2014 17:42:01 (-12.2d)
4	—	nhlx5haze	3.80655	234	Thu, 28 Aug 2014 13:13:45 (-58.2d)
5	—	Opera Solutions	3.78818	187	Thu, 04 Sep 2014 11:46:44 (-3.7h)
6	—	Davut & Josef	3.76856	137	Thu, 04 Sep 2014 18:15:27 (-8.2d)
7	—	Roberto-UCIIM	3.76560	246	Thu, 04 Sep 2014 14:15:52 (-32.6d)
834	↑197	Nemerle	3.25868	51	Thu, 04 Sep 2014 15:06:09 (-16.1h)
835	↓80	CIMFAVuv	3.25090	9	Wed, 23 Jul 2014 16:43:38 (-4.8d)
836	↓80	Jeje	3.25012	4	Sat, 21 Jun 2014 01:11:13
		simple TMVA boosted trees	3.24954		
837	↓80	Xiaohu SUN	3.24954	3	Tue, 03 Jun 2014 13:14:47
838	↓80	Pierre Boutaud	3.24954	10	Fri, 25 Jul 2014 15:25:07 (-30d)

CLASSIFICATION FOR DISCOVERY

#	Δ1w	Team Name <small>‡ model uploaded * in the MONEY</small>	Score	Entries	Last Submission UTC (Best - Last Submission)
1	—	Gábor Melis *	3.85059	97	Thu, 04 Sep 2014 08:07:20 (-34d)
2	↑1	Luboš Motl's team *	3.84522	535	Thu, 04 Sep 2014 09:22:47 (-1.9h)
3	↓1	Tim Salimans *	3.84428	46	Tue, 12 Aug 2014 17:42:01 (-12.2d)
4	—	nhlx5haze	3.80655	234	Thu, 28 Aug 2014 13:13:45 (-58.2d)
5	—	Opera Solutions	3.78818	187	Thu, 04 Sep 2014 11:46:44 (-3.7h)
6	—	Davut & Josef	3.76856	137	Thu, 04 Sep 2014 18:15:27 (-8.2d)
7	—	Roberto-UCIIM	3.76560	246	Thu, 04 Sep 2014 14:15:52 (-32.6d)
834	↑197	Nemerle	3.25868	51	Thu, 04 Sep 2014 15:06:09 (-16.1h)
835	↓80	CIMFAVuv	3.25090	9	Wed, 23 Jul 2014 16:43:38 (-4.8d)
836	↓80	Jeje	3.25012	4	Sat, 21 Jun 2014 01:11:13
		simple TMVA boosted trees	3.24954		
837	↓80	Xiaohu SUN	3.24954	3	Tue, 03 Jun 2014 13:14:47
838	↓80	Pierre Boutaud	3.24954	10	Fri, 25 Jul 2014 15:25:07 (-30d)

CLASSIFICATION FOR DISCOVERY

#	Δ1w	Team Name <small>‡ model uploaded * in the MONEY</small>	Score	Entries	Last Submission UTC (Best - Last Submission)
1	—	Gábor Melis *	3.85059	97	Thu, 04 Sep 2014 08:07:20 (-34d)
2	↑1	Luboš Motl's team *	3.84522	535	Thu, 04 Sep 2014 09:22:47 (-1.9h)
3	↓1	Tim Salimans *	3.84428	46	Tue, 12 Aug 2014 17:42:01 (-12.2d)
4	—	nhlx5haze	3.80655	234	Thu, 28 Aug 2014 13:13:45 (-58.2d)
5	—	Opera Solutions	3.78818	187	Thu, 04 Sep 2014 11:46:44 (-3.7h)
6	—	Davut & Josef	3.76856	137	Thu, 04 Sep 2014 18:15:27 (-8.2d)
7	—	Roberto-UCIIM	3.76560	246	Thu, 04 Sep 2014 14:15:52 (-32.6d)
834	↑197	Nemerle	3.25868	51	Thu, 04 Sep 2014 15:06:09 (-16.1h)
835	↓80	CIMFAVuv	3.25090	9	Wed, 23 Jul 2014 16:43:38 (-4.8d)
836	↓80	Jeje	3.25012	4	Sat, 21 Jun 2014 01:11:13
		simple TMVA boosted trees	3.24954		
837	↓80	Xiaohu SUN	3.24954	3	Tue, 03 Jun 2014 13:14:47
838	↓80	Pierre Boutaud	3.24954	10	Fri, 25 Jul 2014 15:25:07 (-30d)

CLASSIFICATION FOR DISCOVERY

CLASSIFICATION FOR DISCOVERY

- Take home message: running a **data challenge** is a great way to

CLASSIFICATION FOR DISCOVERY

- Take home message: running a **data challenge** is a great way to
 - let **data scientists know** about your problem

CLASSIFICATION FOR DISCOVERY

- Take home message: running a **data challenge** is a great way to
 - let **data scientists know** about your problem
 - get **meaningful solutions** to difficult technical problems

CLASSIFICATION FOR DISCOVERY

- Take home message: running a **data challenge** is a great way to
 - let **data scientists know** about your problem
 - get **meaningful solutions** to difficult technical problems
 - jump-start and nurture **interdisciplinary collaborations**