

# Spectre Attacks: Exploiting Speculative Execution

and why the heck is the computer speculating anyway?



Werner Haas  
April 14, 2021

**CYBERUS**  
TECHNOLOGY

# Spectre Attacks: Exploiting Speculative Execution

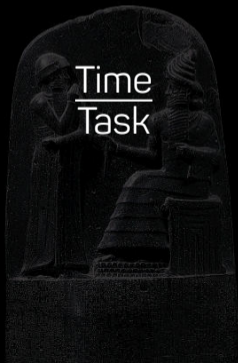
and why the heck is the computer speculating anyway?

Paul Kocher, Jann Horn, Anders Fogh, Daniel Genkin,  
Daniel Gruss, Mike Hamburg, Moritz Lipp, Stefan Mangard,  
Thomas Prescher, Michael Schwarz, Yuval Yarom



**CYBERUS**  
TECHNOLOGY

# Iron Law of Processor Performance



# Iron Law of Processor Performance

The diagram illustrates the Iron Law of Processor Performance equation using four Egyptian reliefs. Each relief depicts a seated figure receiving offerings from a standing figure. The equation is presented as follows:

$$\frac{\text{Time}}{\text{Task}} = \frac{\text{Instructions}}{\text{Task}} \times \frac{\text{Time}}{\text{Cycle}}$$

**Work** **1/Frequency**

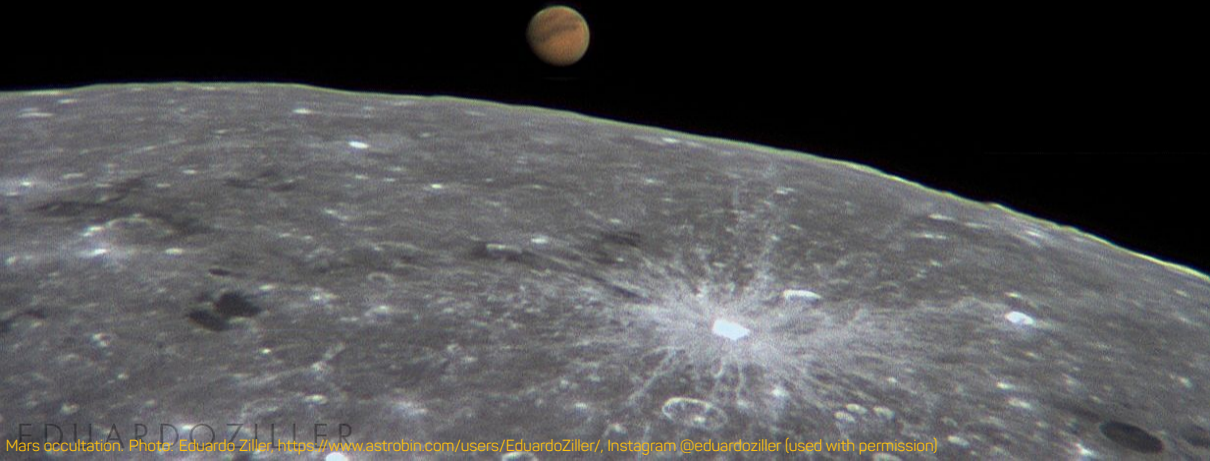
# Iron Law of Processor Performance

The diagram illustrates the Iron Law of Processor Performance equation using four stone tablets. Each tablet features a relief of a seated figure and a standing figure. The equation is presented as follows:

$$\frac{\text{Time}}{\text{Task}} = \frac{\text{Instructions}}{\text{Task}} \times \frac{\text{Cycles}}{\text{Instruction}} \times \frac{\text{Time}}{\text{Cycle}}$$

Below the tablets, the terms are labeled: **Work** (under the second tablet), **CPI** (under the third tablet), and **1/Frequency** (under the fourth tablet).

perf stat md5sum Spectre.pdf



EDUARDO ZILLER

Mars occultation. Photo: Eduardo Ziller, <https://www.astrobin.com/users/EduardoZiller/>, Instagram @eduardoziller (used with permission)

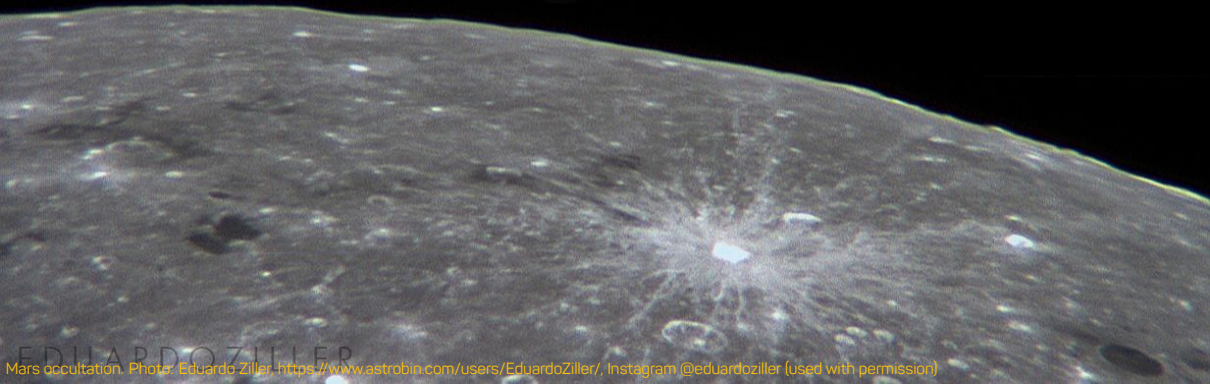
perf stat md5sum Spectre.pdf

≈ 3M cycles

≈ 4M instructions

≈ .5M memory reads

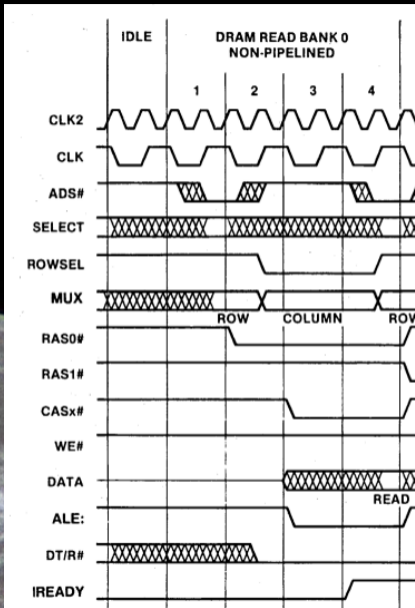
≈ .75 **CPI**





# INTEL 386™ DX MICROPROCESSOR

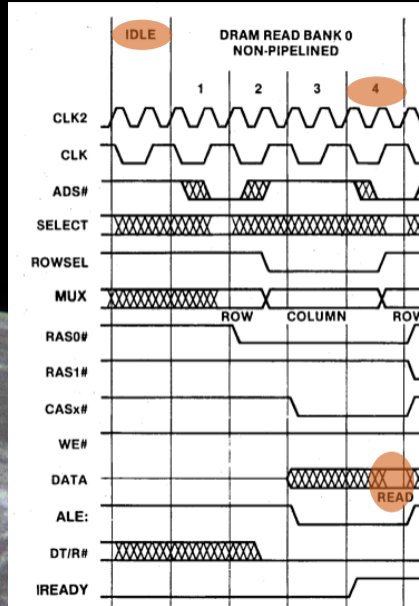
## HARDWARE REFERENCE MANUAL







INTEL 386™ DX MICROPROCESSOR  
HARDWARE REFERENCE MANUAL



# Technology Evolution

Frequency	20 MHz	→	2 GHz
Instructions	4 M		4 M
Memory reads	.5 M		.5 M
Latency [cycles]	4		
Cycles	3 M		
CPI	0,75		

# Technology Evolution

Frequency	20 MHz	→	2 GHz
Instructions	4 M		4 M
Memory reads	.5 M		.5 M
Latency [cycles]	4		>100
Cycles	3 M		
CPI	0,75		

# Technology Evolution

Frequency	20 MHz	→	2 GHz
Instructions	4 M		4 M
Memory reads	.5 M		.5 M
Latency [cycles]	4		>100
Cycles	3 M		>53 M
CPI	0,75		

# Technology Evolution

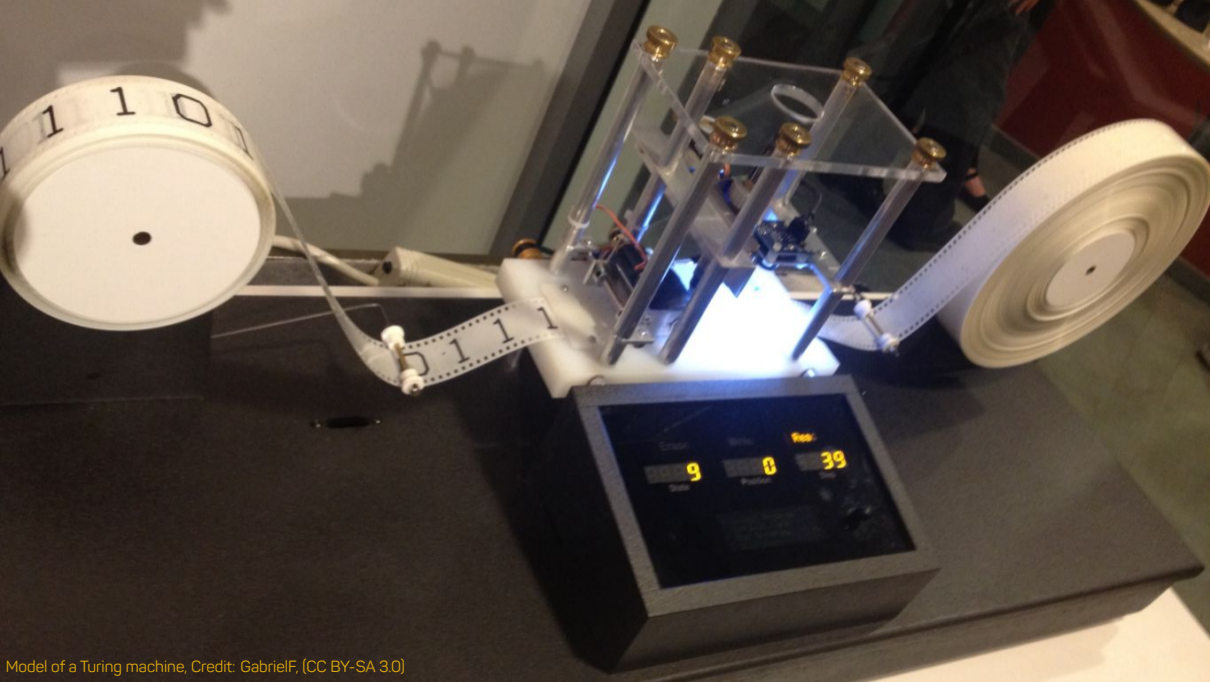
Frequency	20 MHz	→	2 GHz
Instructions	4 M		4 M
Memory reads	.5 M		.5 M
Latency [cycles]	4		>100
Cycles	3 M		>53 M
CPI	0,75	→	>13,25

# Technology Evolution

Frequency	20 MHz	→	2 GHz	100x
Instructions	4 M		4 M	
Memory reads	.5 M		.5 M	
Latency [cycles]	4		>100	
Cycles	3 M		>53 M	
CPI	0,75	→	>13,25	>17x

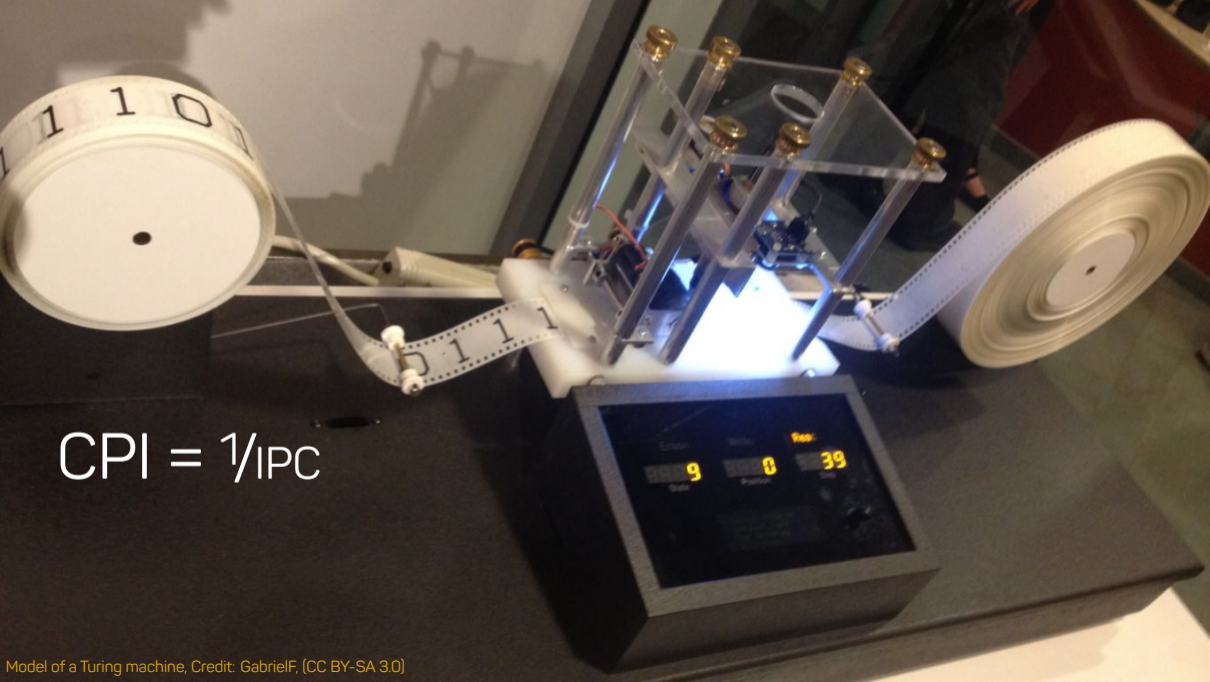
# Technology Evolution



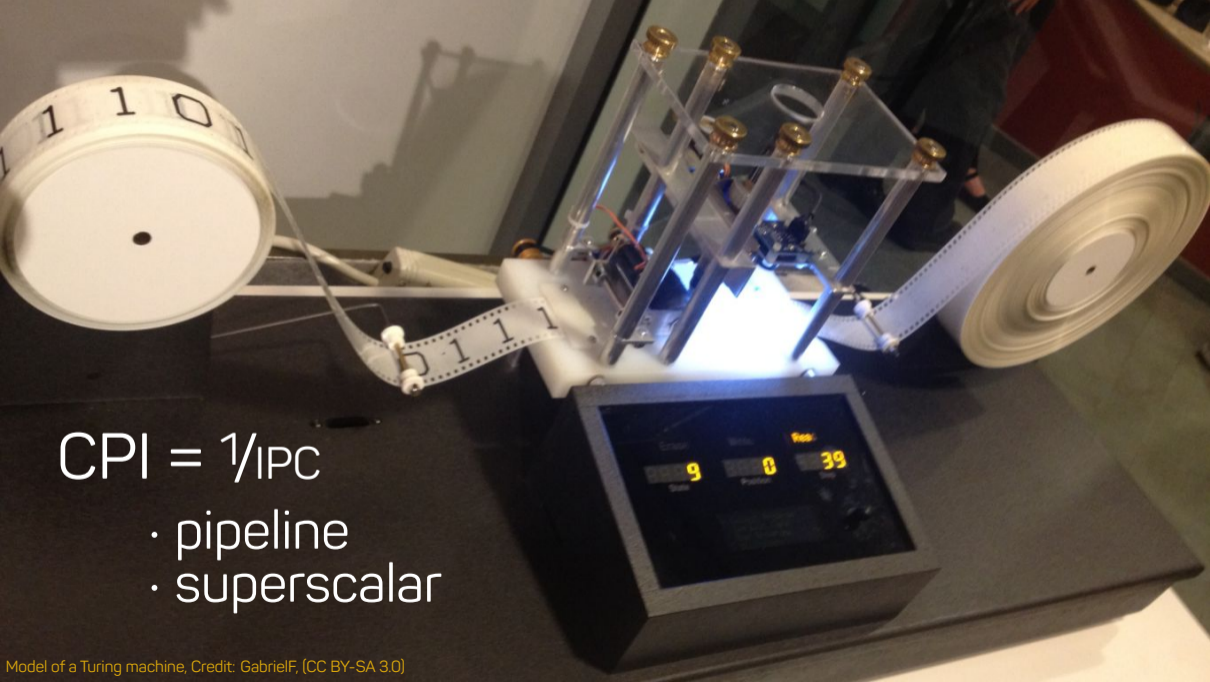


Model of a Turing machine, Credit: GabrielF, (CC BY-SA 3.0)





$$\text{CPI} = 1/\text{IPC}$$



$$CPI = 1/IPC$$

- pipeline
- superscalar

```
# Pad with 0s to get 512b chunks  
while len(message)%512 != 448:  
    message.append(0)
```



```
# Pad with 0s to get 512b chunks  
while len(message)%512 != 448:  
    message.append(0)
```

**jump instr.**  
direct

**detection**  
@decode

**destination**  
calculation

```
# Pad with 0s to get 512b chunks
while len(message)%512 != 448:
    message.append(0)
```

**jump instr.**  
direct  
indirect

**detection**  
@decode  
@decode

**destination**  
calculation  
value lookup

```
# Pad with 0s to get 512b chunks
while len(message)%512 != 448:
    message.append(0)
```

<b>jump instr.</b>	<b>detection</b>	<b>destination</b>
direct	@decode	calculation
indirect	@decode	value lookup
conditional	@execute	calculation

```
# Pad with 0s to get 512b chunks
while len(message)%512 != 448:
    message.append(0)
```

**jump instr.**  
direct  
indirect  
conditional

**detection**  
@decode  
@decode  
@execute

**destination**  
calculation  
value lookup  
calculation

**BHT**

**BTB**

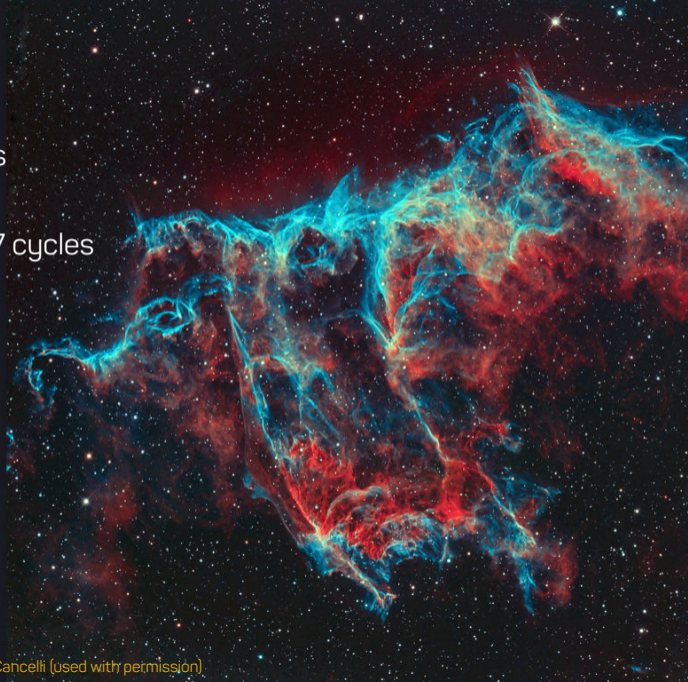
# Spectre





# Spectre v1: BHT

- SPEC CPU 2017:  $\approx 20\%$  branches
- Intel Ice Lake: 5-way decode
- Branch misprediction penalty  $\approx 17$  cycles

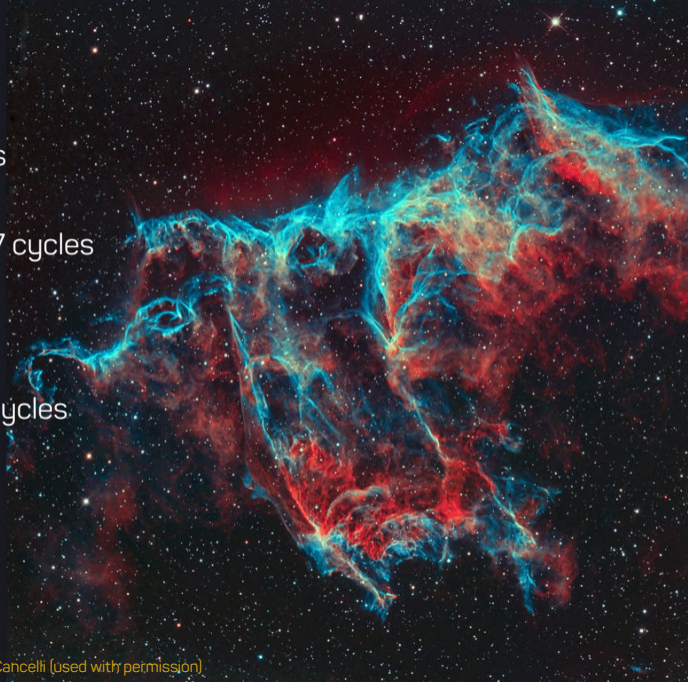


# Spectre v1: BHT

- SPEC CPU 2017:  $\approx 20\%$  branches
- Intel Ice Lake: 5-way decode
- Branch misprediction penalty  $\approx 17$  cycles

Time to fetch 1000 instructions

- perfect prediction:  
 $1000 \text{ instructions} / 5/\text{cycle} = 200 \text{ cycles}$

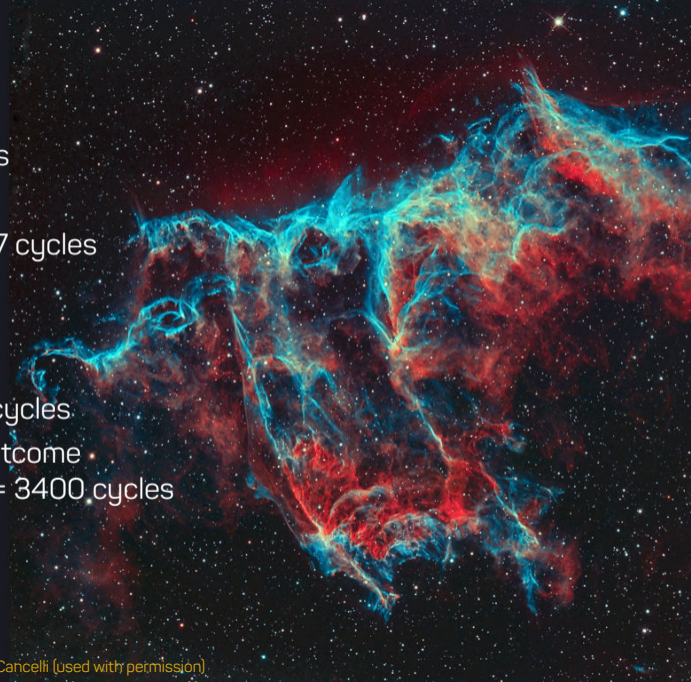


# Spectre v1: BHT

- SPEC CPU 2017:  $\approx 20\%$  branches
- Intel Ice Lake: 5-way decode
- Branch misprediction penalty  $\approx 17$  cycles

Time to fetch 1000 instructions

- perfect prediction:  
 $1000 \text{ instructions} / 5_{\text{cycle}} = 200 \text{ cycles}$
- no speculation: wait for branch outcome  
 $1000 \text{ instructions} / 5 \cdot 17 \text{ cycles} = 3400 \text{ cycles}$

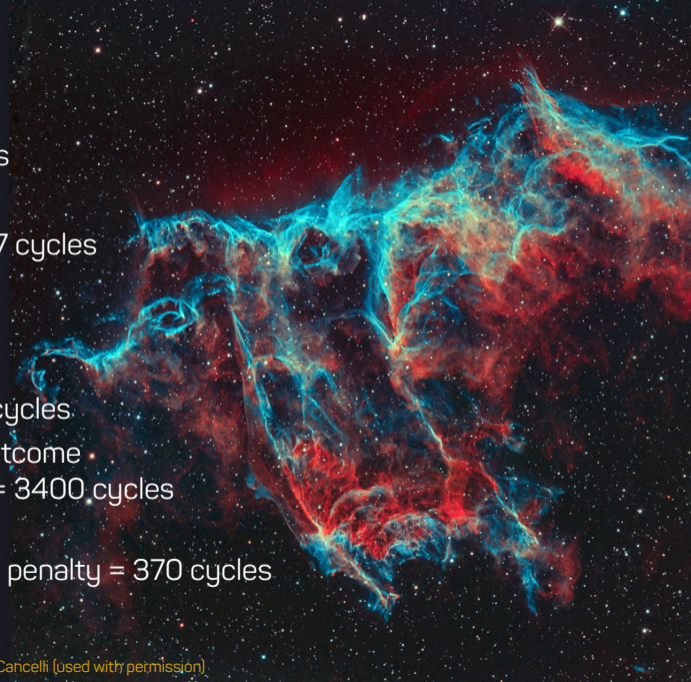


# Spectre v1: BHT

- SPEC CPU 2017:  $\approx 20\%$  branches
- Intel Ice Lake: 5-way decode
- Branch misprediction penalty  $\approx 17$  cycles

## Time to fetch 1000 instructions

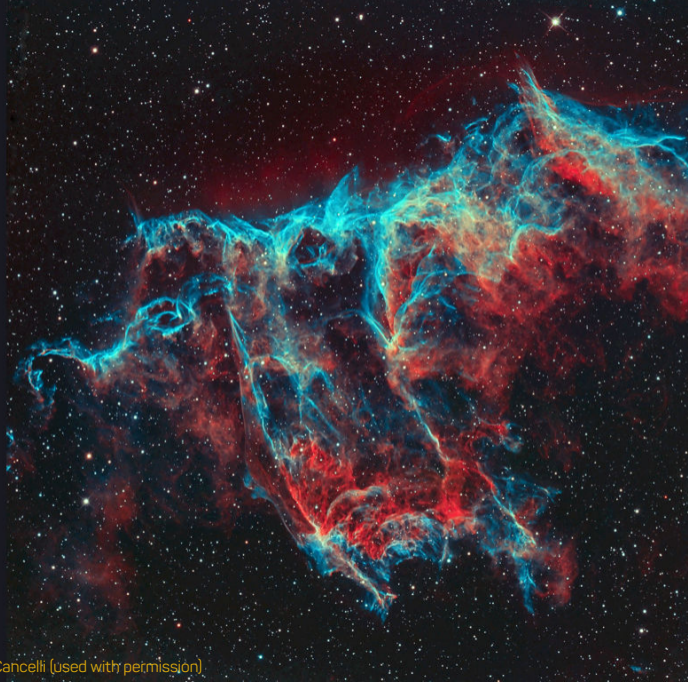
- perfect prediction:  
 $1000 \text{ instructions} / 5_{\text{cycle}} = 200 \text{ cycles}$
- no speculation: wait for branch outcome  
 $1000 \text{ instructions} / 5 \cdot 17 \text{ cycles} = 3400 \text{ cycles}$
- 99% accuracy:  
 $200 \text{ cycles} + 10 \text{ misses} \cdot 17 \text{ cycle penalty} = 370 \text{ cycles}$



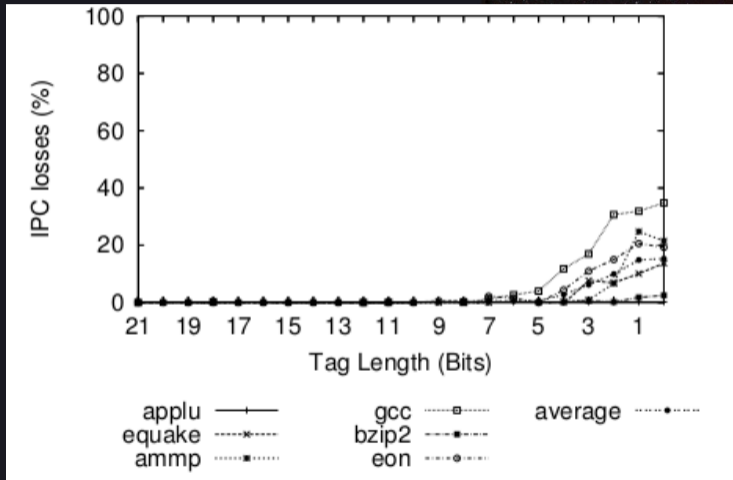
# Spectre v2: BTB

Branch predictor hot spot

- every cycle
- every instruction



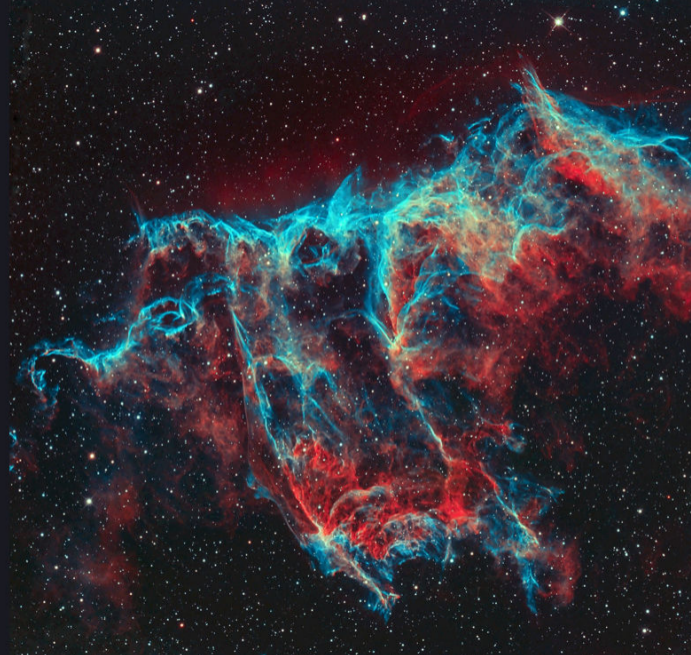
# Spectre v2: BTB



Tomas et al.: Reducing the Number of Bits in the BTB to Attack the Branch Predictor Hot-Spot, Euro-Par 2008



# Spectre







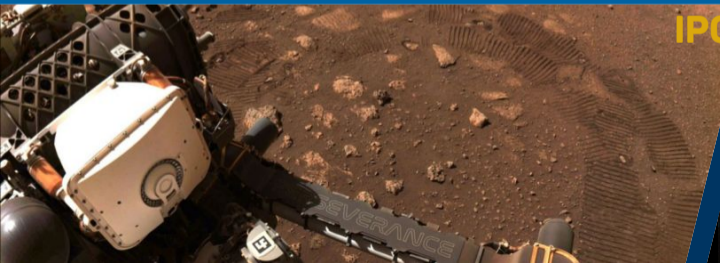
## Iron Law





**Iron Law**  
**CPI**





IPC



Iron Law

CPI



**Spectre**



**Iron Law**



**IPC**

**CPI**

